



**Ecole Doctorale 420
Doctorat de Santé Publique
Spécialité Biostatistiques**

Ludovic TRINQUART

**Impact, détection et correction du biais de publication dans
la méta-analyse en réseau**

Thèse dirigée par le Pr. Gilles CHATELLIER

Soutenue le 28 mars 2013

Jury

Pr. Dominique COSTAGLIOLA, Université Pierre et Marie Curie (Présidente)

Pr. Matthias EGGER, Université de Berne (Rapporteur)

Dr. Raphaël PORCHER, Université Paris Diderot (Rapporteur)

Dr. Jean-Pierre PIGNON, Université Paris Sud (Examineur)

Pr. Gérard BIAU, Université Pierre et Marie Curie (Examineur)

Pr. Gilles CHATELLIER, Université Paris Descartes (Directeur de thèse)

Résumé

La méta-analyse (MA) en réseau, en généralisant la MA conventionnelle, permet d'évaluer toutes les comparaisons deux à deux possibles entre interventions. Les biais de publication ont reçu peu d'attention dans ce contexte.

Nous avons évalué l'impact des biais de publication en utilisant un réseau de 74 essais randomisés évaluant 12 antidépresseurs contre placebo enregistrés à la FDA et un réseau de 51 essais parmi les 74 dont les résultats étaient publiés. Nous avons montré comment les biais de publication biaisaient les quantités d'effet estimées et le classement des traitements. L'effet du biais de publication peut différer entre MA en réseau et MA conventionnelle en ce que les biais affectant un traitement peuvent affecter le classement de tous les traitements.

Nous avons ensuite généralisé un test de détection des biais à la MA en réseau. Il est basé sur la comparaison entre les nombres attendu et observé d'essais avec résultats statistiquement significatifs sur l'ensemble du réseau. Nous avons montré par des études de simulation que le test proposé avait une puissance correcte après ajustement sur l'erreur de type I, excepté lorsque la variance inter-essais était élevée. Par ailleurs, le test indiquait un signal significatif de biais sur le réseau d'essais d'antidépresseurs publiés.

Enfin, nous avons introduit deux modèles d'analyse de sensibilité des résultats d'une MA en réseau aux biais de publication: un modèle de méta-régression qui relie la quantité d'effet estimée à son erreur standard, et un modèle de sélection dans lequel on estime la propension d'un essai à être publié puis l'on redresse le poids des essais en fonction de cette propension. Nous les avons appliqués aux réseaux d'essais d'antidépresseurs.

Ce test et ces modèles d'ajustement tirent leur force de tous les essais du réseau, sous l'hypothèse qu'un biais moyen commun opère sur toutes les branches du réseau.

Mots-clefs

essai contrôlé randomisé; méta-analyse; revue systématique; méta-analyse en réseau; biais de publication; antidépresseurs

Impact, detection and adjustment for reporting bias in network meta-analysis

Abstract

Network meta-analysis (NMA), a generalization of conventional MA, allows for assessing all possible pairwise comparisons between multiple treatments. Reporting bias, a major threat to the validity of MA, has received little attention in the context of NMA.

We assessed the impact of reporting bias empirically using data from 74 FDA-registered placebo-controlled trials of 12 antidepressants and their 51 matching publications. We showed how reporting bias biased NMA-based estimates of treatments efficacy and modified ranking. The effect of reporting bias in NMAs may differ from that in classical meta-analyses in that reporting bias affecting only one drug may affect the ranking of all drugs.

Then, we extended a test to detect reporting bias in network of trials. It compares the number of expected trials with statistically significant results to the observed number of trials with significant p-values across the network. We showed through simulation studies that the test was fairly powerful after adjustment for size, except when between-trial variance was substantial. Besides, it showed evidence of bias in the network of published antidepressant trials.

Finally, we introduced two methods of sensitivity analysis for reporting bias in NMA: a meta-regression model that allows the effect size to depend on its standard error and a selection model that estimates the propensity of trial results being published and in which trials with lower propensity are weighted up in the NMA model. We illustrated their use on the antidepressant datasets.

The proposed test and adjustment models borrow strength from all trials across the network, under the assumption that conventional MAs in the network share a common mean bias mechanism.

Keywords

randomized controlled trials; meta-analysis; systematic review; network meta-analysis; publication bias; small-study effects; antidepressants

Laboratoire de rattachement

INSERM CIC-EC 4, Assistance Publique-Hôpitaux de Paris, Hôpital Européen Georges Pompidou

Liste des productions scientifiques

- Trinquart L, Abbé A, Ravaud P. Impact of reporting bias in network meta-analysis of antidepressant placebo-controlled trials (2012) PLoS ONE, 7 (4):e35219
- Trinquart L, Chatellier G, Ravaud P. Adjustment for reporting bias in network meta-analysis of antidepressant trials. BMC Med Res Methodol. 2012;12(1):150
- Trinquart L, Ioannidis JP, Chatellier G, Ravaud P. A generalized test for reporting bias in trial networks and their meta-analyses: simulation and case studies. Soumis pour publication.

Financement

- Programme Hospitalier de Recherche Clinique National 2011 MIN01-63
- Programme Hospitalier de Recherche Clinique National 2011 MIN01-66
- European Union Seventh Framework Programme (FP7 – HEALTH.2011.4.1-2) n° 285453 (www.open-project.eu)

Remerciements

A Monsieur le Professeur Gilles CHATELLIER, pour la confiance qu'il m'a accordée en dirigeant ce travail. Qu'il trouve ici le témoignage de mes sincères remerciements.

A Monsieur le Professeur Philippe RAVAUD, qu'il trouve ici le témoignage de ma reconnaissance et de mes sincères remerciements.

A Monsieur le Professeur Matthias EGGER, à Monsieur le Docteur Raphaël PORCHER, qui m'ont fait l'honneur d'accepter d'être rapporteurs et membres du jury. Qu'ils trouvent ici le témoignage de ma profonde considération et de mes sincères remerciements.

A Madame le Professeur Dominique COSTAGLIOLA, à Monsieur le Docteur Jean-Pierre PIGNON, à Monsieur le Professeur Gérard BIAU, qui m'ont fait l'honneur d'accepter d'être membres du jury. En témoignage de mes sentiments respectueux et de ma sincère considération.

Table des matières

1. INTRODUCTION	8
1.1 De l'essai randomisé à la méta-analyse.....	8
1.2 Biais de publication dans la méta-analyse.....	12
1.3 De la méta-analyse à la méta-analyse en réseau	18
1.4 Biais de publication et méta-analyse en réseau.....	20
2. METHODES STATISTIQUES DE LA META-ANALYSE EN RESEAU	22
2.1 Comparaisons indirectes ajustées	22
2.2 Méta-analyse en réseau – modèle à effet fixe	23
2.3 Méta-analyse en réseau – modèle à effets aléatoires	25
2.4 Méta-analyse en réseau – paramétrisation de Lu et Ades	26
2.5 Extension du modèle aux essais à plus de deux bras	26
2.6 Présentation des résultats d'une méta-analyse en réseau	29
2.7 Approche bayésienne et approche fréquentiste	31
2.8 Hypothèses nécessaires.....	32
2.9 Evaluation de l'incohérence et modèles d'incohérence	33
2.9.1 Méthode de Bucher.....	33
2.9.2 Modèle de Lumley	34
2.9.3 Modèle de Lu et Ades	35
2.9.4 Essais à plus de deux bras et incohérence	37
3. IMPACT DU BIAIS DE PUBLICATION DANS LA META-ANALYSE EN RESEAU	38
3.1 Contexte.....	38
3.2 Méthodes	40
3.2.1 Données issues de la FDA et des articles publiés	40
3.2.2 Méta-analyses en réseau	41
3.2.3 Impact des biais de publication.....	42
3.2.4 Impact du biais de publication affectant un seul médicament.....	43
3.2.5 Impact spécifique du biais de publication et de la présentation sélective des analyses.....	44
3.3 Résultats	45
3.3.1 Impact des biais de publication.....	47
3.3.2 Impact du biais de publication affectant un seul médicament.....	49
3.3.3 Impact spécifique du biais de publication et de la présentation sélective des analyses.....	51
3.4 Discussion	52
3.5 Conclusion.....	58
4. DETECTION DU BIAIS DE PUBLICATION DANS LA META-ANALYSE EN RESEAU	59
4.1 Contexte.....	59
4.2 Test d'excès de résultats significatifs	62
4.2.1 Détection du biais dans une méta-analyse conventionnelle	62
4.2.2 Détection du biais dans une méta-analyse en réseau.....	63

4.3. Simulation	64
4.3.1 Méthodes pour générer les données d'un réseau d'essais	65
4.3.2 Scénarios étudiés	71
4.3.3 Critères d'évaluation de la performance du test	72
4.3.4 Nombre de simulations à effectuer	74
4.4 Résultats	74
4.4.1 Méta-analyse conventionnelle	74
4.4.2 Méta-analyse en réseau	78
4.5 Application du test à deux réseaux d'essais	81
4.5.1 Réseau d'essais d'antidépresseurs	81
4.5.2 Réseau d'essais d'antipsychotiques	83
4.6 Discussion	84
4.7 Conclusion	88
 5. CORRECTION DU BIAIS DE PUBLICATION DANS LA META-ANALYSE EN RESEAU	 90
 5.1 Contexte	 90
5.2 Méthodes	91
5.2.1 Bases de données utilisées	92
5.2.2 Méta-analyse en réseau	93
5.2.3 Modèles d'ajustement	94
5.2.4 Estimation des modèles	97
5.2.5 Comparaison des modèles	98
5.3 Résultats	99
5.4 Discussion	106
5.5 Conclusion	111
 6. CONCLUSION	 112
 BIBLIOGRAPHIE	 116
 ANNEXES	 123

Chapitre 1

Introduction

1.1 De l'essai randomisé à la méta-analyse

1.2 Biais de publication dans la méta-analyse

1.3 De la méta-analyse à la méta-analyse en réseau

1.4 Biais de publication et méta-analyse en réseau

1.1 De l'essai randomisé à la méta-analyse

L'évaluation thérapeutique repose principalement, depuis les années 1950, sur l'essai contrôlé randomisé (Chalmers 2010). Ces essais ont été conçus et interprétés pendant longtemps de manière isolée. L'absence d'une vue d'ensemble, pour la plupart des interventions en santé, des informations nécessaires à leur évaluation rigoureuse, a été mise en évidence par Archibald Cochrane dans les années 1970: « Une grande critique de notre profession est cette incapacité à établir un résumé, discipline par discipline, révisé périodiquement, de tous les essais contrôlés randomisés publiés sur un sujet donné » (Cochrane 1972).

Un changement de paradigme a donc consisté à examiner l'ensemble de tous les essais randomisés qui ont abordé une même question de recherche. Cet examen doit se faire dans le cadre d'une revue systématique qui correspond à l'approche scientifique rigoureuse consistant à rechercher, sélectionner et évaluer tous les essais pertinents de manière transparente et reproductible (Egger, Davey Smith et al. 2001). Une revue systématique évalue généralement l'efficacité et la sécurité d'une intervention expérimentale donnée par rapport à un comparateur donné (par exemple, un placebo). En outre, les résultats des essais peuvent être synthétisés à l'aide d'une méta-analyse, ce qui permet de surmonter les limites des essais individuels en

réduisant l'incertitude dans l'effet du traitement combiné et en offrant la possibilité d'explorer la variabilité des effets du traitement d'un essai à l'autre (hétérogénéité). La méta-analyse est réalisée la plupart du temps sur les données agrégées de chaque essai, extraites d'articles rapportant leurs résultats, bien qu'il existe un nombre croissant - une cinquantaine par an de 2005 à 2009 - de méta-analyses réalisées à partir des données individuelles des patients (en anglais, individual patient data meta-analysis) (Riley, Lambert et al. 2010; Ahmed, Sutton et al. 2012).

Le principe de la méthode statistique consiste à choisir une mesure de l'effet du traitement, à estimer l'effet du traitement expérimental (relativement au témoin) dans chaque essai ainsi que la variance associée, puis à calculer une moyenne pondérée de ces estimations. Intuitivement, on souhaite accorder plus de poids aux essais qui apportent plus d'information, c'est-à-dire ceux pour lesquels la variance sur l'estimation de l'effet du traitement est la plus petite (Borenstein, Hedges et al. 2009).

Admettons que pour chaque essai i , on observe r_{ij} et r_{ik} événements parmi n_{ij} et n_{ik} patients randomisés dans le groupe expérimental j et le groupe témoin k , respectivement. Sur une échelle relative, on peut estimer le rapport des cotes d'événement entre j et k par $\frac{r_{ij}}{n_{ij}-r_{ij}} / \frac{r_{ik}}{n_{ik}-r_{ik}}$, et la variance du logarithme du rapport des cotes par $\frac{1}{r_{ij}} + \frac{1}{n_{ij}-r_{ij}} + \frac{1}{r_{ik}} + \frac{1}{n_{ik}-r_{ik}}$; on peut également estimer le risque relatif d'événement entre j et k par $\frac{r_{ij}}{n_{ij}} / \frac{r_{ik}}{n_{ik}}$ et la variance du logarithme du risque relatif par $\frac{1}{r_{ij}} - \frac{1}{n_{ij}} + \frac{1}{r_{ik}} - \frac{1}{n_{ik}}$ (Fleiss, Levin et al. 2004).

Pour un critère de jugement continu, admettons que nous observions les moyennes et les écarts types m_{ij} , s_{ij} et m_{ik} , s_{ik} sur n_{ij} et n_{ik} patients dans les groupes expérimental et témoin de l'essai i , respectivement. On peut estimer la différence de moyennes entre j et k par $m_{ij} - m_{ik}$ et la variance commune associée par $s_i^2 = [(n_{ij} - 1)s_{ij}^2 + (n_{ik} - 1)s_{ik}^2] / (n_{ij} + n_{ik} - 2)$. Si

différentes échelles de mesure ont été utilisées pour mesurer le critère d'un essai à l'autre, on peut estimer la différence de moyennes standardisée entre j et k par la mesure d de Cohen : $d_i = (m_{ij} - m_{ik})/s_i$. La variance associée est estimée par $(n_{ij} + n_{ik})/(n_{ij} \cdot n_{ik}) + d_i^2/[2 \cdot (n_{ij} + n_{ik})]$. Cet estimateur est biaisé sur les échantillons de petites tailles et, le cas échéant, on peut appliquer un facteur de correction multiplicatif pour estimer ainsi la mesure g de Hedges : $g_i = J(n_{ij} + n_{ik} - 2) \cdot d_i$, avec $J(a) = \Gamma(a)/[\sqrt{a/2} \cdot \Gamma((a-1)/2)]$. On peut montrer que $J(n_{ij} + n_{ik} - 2) \approx (1 - 3/[4(n_{ij} + n_{ik}) - 9])$. La variance associée est estimée par $J(n_{ij} + n_{ik} - 2)^2 \cdot var(d_i)$ (Hedges and Olkin 1985).

Après avoir calculé une estimation $\hat{\theta}_i$ de l'effet du traitement et sa variance associée v_i pour chacun des N essais sélectionnés, l'estimation de l'effet combiné par une méta-analyse à effet fixe est donnée par $\hat{\theta}_{fixe} = \sum_{i=1}^N w_i \cdot \hat{\theta}_i / \sum_{i=1}^N w_i$ avec $w_i = 1/v_i$. La variance associée est estimée par $var(\hat{\theta}_{fixe}) = 1/\sum_{i=1}^N w_i$.

L'hétérogénéité statistique, c'est-à-dire la variabilité vraie des effets du traitement d'un essai à l'autre, peut être évaluée par l'exploration visuelle des données et plusieurs représentations graphiques sont disponibles : le graphique en forêt (en anglais, forest plot), le graphique de Galbraith (Galbraith 1988), le graphique de l'Abbé (L'Abbe, Detsky et al. 1987), le graphique de Baujat (Baujat, Mahe et al. 2002). On peut estimer la statistique $Q = \sum_{i=1}^N w_i \cdot (\hat{\theta}_i - \hat{\theta}_{fixe})^2$; elle est distribuée selon une distribution χ^2 à $N-1$ degrés de liberté de sorte que l'on peut dériver un test de l'hypothèse d'homogénéité. On estime également le coefficient I^2 de Higgins : $100\% \times (Q - N + 1) / Q$ (Higgins and Thompson 2002; Higgins, Thompson et al. 2003). Les valeurs négatives de I^2 sont ramenées à 0%, de sorte que I^2 est compris entre 0% et 100%.

Une valeur de 0% indique l'absence d'hétérogénéité, et des valeurs plus élevées indiquent une hétérogénéité croissante. Enfin, on peut mesurer l'hétérogénéité statistique en estimant directement la variance inter-essais τ^2 de l'effet du traitement. Par exemple, l'estimateur de DerSimonian et Laird par la méthode des moments s'écrit : $\hat{\tau}^2 = \frac{Q - N + 1}{\sum_{i=1}^N w_i - \left[\sum_{i=1}^N w_i \right]^2 / \sum_{i=1}^N w_i^2}$ (DerSimonian and Laird 1986). Cet estimateur est le plus fréquemment utilisé en pratique, mais d'autres estimateurs plus performants sont disponibles (Sidik and Jonkman 2005; Sidik and Jonkman 2007).

Une manière d'incorporer l'hétérogénéité dans la synthèse est d'estimer l'effet combiné par une méta-analyse à effets aléatoires : $\hat{\theta}_{aléatoire} = \sum_{i=1}^N w_i^* \cdot \hat{\theta}_i / \sum_{i=1}^N w_i^*$ avec $w_i^* = 1/(v_i + \tau^2)$. La variance associée est estimée par $var(\hat{\theta}_{aléatoire}) = 1 / \sum_{i=1}^N w_i^*$. Dans le modèle à effet fixe, on suppose qu'il existe un effet vrai du traitement commun sous-jacent à tous les essais et les estimations de l'effet du traitement dans chaque essai s'écartent de cet effet vrai à cause des fluctuations d'échantillonnage uniquement : $\hat{\theta}_i \sim N(\theta, v_i)$. Dans le modèle à effets aléatoires, on suppose qu'il existe un effet vrai sous-jacent à chaque essai (l'estimation de l'effet dans chaque essai s'écarte de cet effet vrai à cause des fluctuations d'échantillonnage) et on suppose que les effets vrais sont distribués selon une loi normale de moyenne θ et de variance τ^2 : $\hat{\theta}_i \sim N(\theta_i, v_i)$ et $\theta_i \sim N(\theta, \tau^2)$ (Riley, Higgins et al. 2011).

Rappelons enfin que la méta-analyse consiste également à explorer les sources d'hétérogénéité parmi les facteurs de diversité clinique ou méthodologique. On peut utiliser l'analyse en sous-groupes, où l'on évalue l'hétérogénéité et l'effet combiné au sein de strates d'essais, ou la méta-régression, où l'on modélise la relation entre l'effet traitement et une covariable x_i : $\hat{\theta}_i \sim N(\theta_i + \beta x_i, v_i)$ et $\theta_i \sim N(\theta, \tau^2)$ (Thompson and Sharp 1999; Thompson and Higgins 2002).

1.2 Biais de publication dans la méta-analyse

Il existe deux principales menaces à la validité des résultats d'une méta-analyse : les biais liés aux études individuelles et les méta-biais.

Rappelons brièvement que le premier type de biais concerne des procédés relatifs aux études individuelles: par exemple, l'allocation par tirage au sort des traitements évalués aux patients dans un essai aide à prévenir les biais de sélection. Un essai dans lequel cette procédure de randomisation n'est pas adéquate est à risque de biais, c'est-à-dire que l'estimation de l'effet traitement fournie par cet essai et, par conséquent, l'estimation combinée fournie par la méta-analyse, risquent d'être biaisées (en anglais, garbage in, garbage out) (Higgins, Altman et al. 2011). Des études méta-épidémiologiques, en examinant de nombreuses méta-analyses sur divers sujets, ont permis d'identifier les facteurs de risque de biais dans les méta-analyses, qui ont trait à la fois aux défauts méthodologiques et à des caractéristiques des études comme la nature monocentrique (Dechartres, Boutron et al. 2011; Savovic, Jones et al. 2012).

Les méta-biais ne sont, eux, pas liés à des procédés relatifs aux essais individuels, mais se manifestent seulement lorsque l'on examine une collection d'essais (Goodman and Dickersin 2011). Le biais de publication (en anglais, publication bias) est un méta-biais : il s'agit du fait de ne pas disposer dans une méta-analyse des résultats d'une certaine proportion des études existantes en raison de leurs résultats négatifs (Montori, Ioannidis et al. 2008; Dickersin and Chalmers 2011). L'existence du biais de publication a été maintes fois documentée dans des études montrant que les essais avec des résultats « positifs » - c'est-à-dire favorisant le traitement expérimental, statistiquement significatifs - sont publiés plus souvent, plus rapidement, plus facilement dans des revues à facteur d'impact élevé, et plus fréquemment dans des revues de langue anglaise, que ne le sont les essais avec des résultats « négatifs » - ne favorisant pas le

traitement expérimental, ne montrant pas de différence statistiquement significative (Hopewell, Loudon et al. 2009; Song, Parekh-Bhurke et al. 2009). Si les résultats « positifs » sont plus susceptibles d'être disponibles, une méta-analyse basée sur ces résultats sera biaisée (Song, Parekh et al. 2010).

On sait désormais que le biais de publication fait partie d'une classe de biais qui englobe un large éventail de comportements liés à la présentation des résultats (en anglais, reporting bias). En outre, la présentation sélective de certaines analyses statistiques et de certains critères de jugement (en anglais, selective analysis and outcome reporting bias) est de plus en plus considérée comme une menace potentiellement plus grande que le biais de publication spécifiquement (Chan, Hrobjartsson et al. 2004; Dwan, Altman et al. 2008; Mathieu, Boutron et al. 2009; Smyth, Kirkham et al. 2011). Dans la suite du manuscrit, nous avons préféré conserver l'expression « biais de publication » pour désigner l'ensemble de la classe des biais plutôt que d'introduire des expressions résultant de la traduction littérale des termes anglais (reporting bias, dissemination bias). Si nécessaire, nous avons parlé du « biais de publication spécifiquement » pour désigner la suppression d'un essai dans son intégralité.

Bien que les chercheurs et les éditeurs de revues médicales soient eux-mêmes responsables de ces pratiques, il faut souligner que la majorité des exemples de présentation sélective d'essais entiers ou de résultats (analyses statistiques ou critères de jugement) sont liés à la recherche promue et financée par l'industrie pharmaceutique (Lexchin 1993; Lexchin, Bero et al. 2003; Gotzsche 2005; Gotzsche, Hrobjartsson et al. 2007; Gotzsche 2012; Lexchin 2012). On peut même considérer que la promotion ou le financement des essais par l'industrie pharmaceutique constituent un méta-biais (Lundh, Sismondo et al. 2012). Les études sur un médicament ou un dispositif médical promues et financées par le fabricant montrent, en moyenne, des résultats plus favorables (par exemple, des valeurs p significatives) que les études

promues par d'autres sources. Or cette association - résultats plus favorables dans les études promues par l'industrie pharmaceutique - est expliquée par des facteurs autres que ceux documentés par l'outil Cochrane d'évaluation du «risque de biais». La promotion et le financement par l'industrie ne sont pas en soi un processus générateur de biais - comme l'est par exemple l'assignation secrète inadéquate - mais un facteur de risque de biais. Il existe de nombreux mécanismes subtils par lesquels la promotion et le financement par l'industrie peuvent influencer les résultats, et l'évaluation de cette caractéristique devrait être utilisée comme variable de substitution de ces mécanismes.

Les solutions à ces problèmes de biais de publication viendront de l'amélioration de la transparence de la recherche (Zarin and Tse 2008). Une première étape a été franchie avec les registres d'essais cliniques (U.S. National Institutes of Health ; World Health Organization ; Dickersin and Rennie 2003). Les promoteurs d'essais ont l'obligation d'enregistrer les essais et un protocole abrégé avant l'inclusion du premier patient, de sorte que les méta-analystes peuvent détecter des essais non publiés ou des écarts par rapport au plan préétabli pour la conduite de l'essai ou de l'analyse (De Angelis, Drazen et al. 2004). L'étape suivante consistera en un accès aux protocoles d'essais dans leur intégralité et en un accès facilité aux données d'essais, d'abord aux données agrégées (enregistrées auprès des régulateurs ou sur les registres d'essais) puis directement aux données individuelles (Hrynaskiewicz and Altman 2009; Gotzsche 2011; Zarin, Tse et al. 2011; Godlee and Groves 2012).

Du point de vue statistique, l'objectif est de faire le meilleur usage possible des données à disposition dans la méta-analyse. On dispose d'outils graphiques et de tests statistiques pour évaluer la présence de biais et de méthodes d'analyse de sensibilité des résultats aux biais (Rothstein, Sutton et al. 2005).

L'outil le plus fréquemment utilisé est le graphique en entonnoir (en anglais, funnel plot) (Egger, Davey Smith et al. 1997; Sterne, Egger et al. 2001). Il représente, pour chaque essai, une mesure de la taille de l'essai en ordonnée par rapport à l'estimé ponctuel de l'effet du traitement en abscisse. La précision de l'estimation de l'effet vrai du traitement augmente avec la taille des essais: les estimés ponctuels dont la précision est faible (essais de petite taille) varient autour de l'effet vrai avec une plus grande amplitude que ceux dont la précision est grande (essais de grande taille). En l'absence de biais, les estimés ponctuels sont répartis de façon homogène et symétrique autour de l'effet vrai du traitement. Le nuage de points a la forme d'un entonnoir renversé. En cas de biais, par exemple si les études de petite taille ne montrant pas d'effet significatif ne sont pas publiées, le nuage de points sera asymétrique. Plus l'asymétrie est prononcée, plus la suspicion de biais est forte. Il faut garder présent à l'esprit que le biais de publication n'est qu'une des causes possibles de l'asymétrie. D'autres causes possibles incluent les défauts méthodologiques, qui conduisent à une surestimation de l'effet du traitement et qui sont plus fréquents dans les essais de petite taille, ou l'hétérogénéité statistique. Par conséquent, ce graphique reste avant tout un moyen d'examiner l'effet des études de petite taille (en anglais, small-study effect), la tendance des études de petite taille à montrer des estimations plus grandes de l'effet du traitement (Nuesch, Trelle et al. 2010; Sterne, Sutton et al. 2011). Malgré des améliorations récentes, l'outil présente l'inconvénient majeur de reposer sur une interprétation visuelle subjective du degré d'asymétrie (Terrin, Schmid et al. 2005; Peters, Sutton et al. 2008; Langan, Higgins et al. 2012). Le graphique reste peu utile dans la majorité des méta-analyses qui comportent dans trois quarts des cas moins de 6 essais (Davey, Turner et al. 2011).

De nombreux tests statistiques ont été proposés pour détecter la présence de biais dans une méta-analyse (Sterne, Sutton et al. 2011). La plupart d'entre eux évaluent formellement le degré d'asymétrie dans le graphique en entonnoir comme un signe de l'effet des études de petite

taille, c'est-à-dire la tendance qu'ont les petits essais à montrer des estimations de l'effet du traitement plus grandes, quelles que soient les raisons de cet effet (Begg and Mazumdar 1994; Egger, Davey Smith et al. 1997; Macaskill, Walter et al. 2001; Schwarzer, Antes et al. 2002; Harbord, Egger et al. 2006; Peters, Sutton et al. 2006; Schwarzer, Antes et al. 2007; Rucker, Schwarzer et al. 2008). Entre autres, le test d'Egger est basé sur une régression linéaire entre l'estimation de l'effet du traitement et l'erreur standard associée, pondérée par l'inverse de la variance de l'estimation de l'effet du traitement. Ces tests ne pallient pas en pratique aux défauts du graphique en entonnoir. D'une part, ils sont généralement peu puissants, de sorte que même si un test ne fournit pas la preuve de l'asymétrie, le biais ne peut être exclu. D'autre part, les critères d'applicabilité de ces tests (absence d'hétérogénéité statistique importante, $P < 50\%$; au moins 10 études dont une montrant des résultats statistiquement significatifs; rapport entre la variance maximale et la variance minimale supérieure à 4) sont rarement vérifiés dans les méta-analyses (Ioannidis and Trikalinos 2007).

Enfin, plusieurs méthodes d'analyse de sensibilité des résultats aux biais ont été introduites (Sutton, Song et al. 2000). On distingue les modèles de sélection, la méthode dite « supprimer et remplacer » (en anglais, trim and fill), et des méthodes basées sur la méta-régression. Brièvement, dans les modèles de sélection, des fonctions de pondération sont utilisées pour modéliser le processus de sélection des essais (Iyengar and Greenhouse 1988; Hedges 1992; Vevea and Hedges 1995; Givens, Smith et al. 1997; Silliman 1997; Vevea and Woods 2005; Rufibach 2011). Ces modèles sont basés sur un modèle spécifiant la distribution des estimations de l'effet du traitement en l'absence de processus de sélection, et un modèle de sélection qui spécifie comment cette distribution est modifiée par le processus de sélection. On suppose que la probabilité de sélectionner un essai dans la méta-analyse est liée à la significativité statistique de ses résultats ou, dans le modèle de Copas, à la fois à la taille de l'essai et à l'intensité de

l'estimation de l'effet du traitement (Copas and Shi 2000; Copas and Shi 2001; Carpenter, Schwarzer et al. 2009; Schwarzer, Carpenter et al. 2010). La méthode « trim and fill » est une méthode non paramétrique qui consiste à supprimer, de manière itérative, les études entraînant l'asymétrie du graphique en entonnoir (les petites études montrant les résultats les plus favorables), jusqu'à ce que le graphique en entonnoir sur les points restants permette d'estimer l'axe de symétrie vrai du graphique, puis à replacer les études omises et à ajouter les points symétriques par rapport à cet axe de symétrie (les études « manquantes ») (Duval and Tweedie 2000; Sutton, Duval et al. 2000). La méthode fournit donc l'estimation du nombre d'essais manquants et d'un effet ajusté en effectuant une méta-analyse incluant les études ajoutées. Enfin, des méthodes basées sur la méta-régression ont été introduites plus récemment. Le modèle décrit par Moreno et al suppose, à la manière du test d'Egger, une relation linéaire entre la quantité d'effet estimée et son écart type ou sa variance, et implique une extrapolation au-delà des données observées à une étude hypothétique de précision infinie (Moreno, Sutton et al. 2009; Moreno, Sutton et al. 2012). La méta-analyse limite décrite par Rucker et al repose sur le même principe (Rucker, Carpenter et al. 2011; Rucker, Schwarzer et al. 2011). Toutes ces méthodes doivent être considérées comme analyses de sensibilité et non comme fournissant des estimations définitives corrigées du biais de publication. Si les résultats d'une méta-analyse changent après avoir utilisé une des méthodes décrite, alors il faut interpréter la méta-analyse avec prudence. Enfin, il convient de remarquer que la méta-analyse à effets aléatoires sera plus affectée par les effets du biais de publication. En effet, le modèle à effets aléatoires donne plus de poids relativement au modèle à effet fixe aux études de petite taille, qui sont plus susceptibles d'être supprimées. De plus, le biais de publication aura également un effet sur l'estimation de la variance inter-essais, ce qui affectera à son tour l'estimé combiné du modèle à effets aléatoires (Jackson 2007).

1.3 De la méta-analyse à la méta-analyse en réseau

Un deuxième changement de paradigme est en cours. En effet, pour beaucoup de maladies, plusieurs interventions thérapeutiques sont disponibles (par exemple, il y a 68 antidépresseurs au choix) et la nouvelle question posée par de nombreux professionnels de santé est la suivante: quelle est la meilleure intervention parmi les différentes interventions existantes pour une condition spécifique? Comme la prise de décision nécessiterait idéalement des comparaisons directes entre toutes les interventions, les chercheurs examinent à présent des réseaux d'essais, c'est-à-dire tous les essais randomisés comparant les différentes interventions disponibles pour traiter une maladie donnée, les uns contre les autres ou contre d'autres comparateurs (par exemple, un placebo, pas de traitement, d'autres traitements actifs) (Salanti, Higgins et al. 2008; Salanti, Kavvoura et al. 2008).

Ces réseaux d'essais sont généralement représentés par un graphe non orienté, dans lequel les nœuds représentent les interventions et deux nœuds sont reliés par une ligne si au moins un essai contrôlé randomisé a comparé les deux interventions correspondantes. L'épaisseur des lignes est proportionnelle au nombre d'essais. La taille d'un nœud est proportionnelle au nombre de sujets qui ont reçu l'intervention correspondante. Le graphique est généralement connecté, c'est-à-dire qu'il doit y avoir une chaîne de comparaisons directes qui relie chaque intervention avec d'autres interventions. Cependant, certaines comparaisons directes peuvent être manquantes.

Il en résulte des schémas de comparaisons complexes, similaires à la structure des plans d'expérience en blocs incomplets. En effet, souvent, les nombreuses approches thérapeutiques pour une maladie donnée n'ont jamais été comparées par des essais randomisés face-face. Il y a souvent une prédilection pour les essais contrôlés par placebo (en raison des promoteurs ou des

agences de réglementation) (Lathyris, Patsopoulos et al. 2010; Goldberg, Schneeweiss et al. 2011; Estellat and Ravaud 2012).

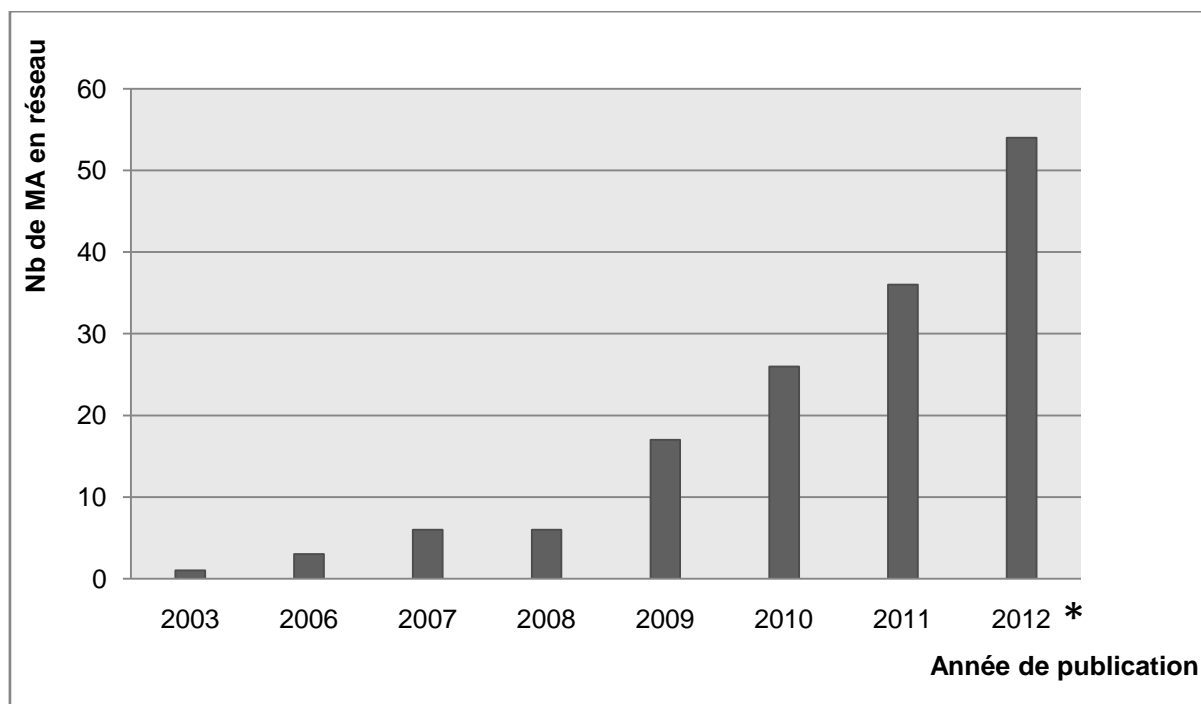
Il peut alors être très informatif d'examiner la structure des comparaisons entre les différentes interventions dans un réseau. En effet, la géométrie du réseau reflète éventuellement la façon dont le programme de recherche dans son ensemble a été construit. L'évolution du réseau a pu être influencée par des biais spécifiques, c'est-à-dire en évitant certaines comparaisons et en en préférant d'autres, notamment les comparaisons contre placebo. La géométrie de ces réseaux peut être résumée par plusieurs métriques, mesurant la diversité, la cooccurrence et l'homophilie (Salanti, Higgins et al. 2008; Rizos, Salanti et al. 2011). Plusieurs exemples, dans le domaine des antifongiques, des stratégies thérapeutiques du cancer avancé du sein, du myélome multiple ou de la polyarthrite rhumatoïde, ont montré par l'examen de la géométrie du réseau comment certaines comparaisons ont été évitées, conduisant à un agenda de recherche biaisé (Mauri, Polyzos et al. 2008; Georgiev, Tsalatsanis et al. 2011; Rizos, Salanti et al. 2011; Estellat and Ravaud 2012).

Au-delà de l'examen de la géométrie du réseau, on peut utiliser le réseau pour synthétiser toutes les comparaisons. Une telle synthèse devrait permettre d'évaluer l'efficacité relative et/ou la sécurité relative de toutes les interventions disponibles. Or les méta-analyses conventionnelles sont limitées à cet égard puisqu'elles portent sur la synthèse des essais comparant deux interventions spécifiques. Cependant, une méthode de synthèse plus récente, la méta-analyse en réseau (en anglais, network meta-analysis, multiple-treatment meta-analysis, ou mixed treatment comparison meta-analysis), permet d'estimer toutes les comparaisons deux à deux entre les interventions, même si certaines n'ont jamais été traitées par aucun essai.

Ces dernières années ont vu une augmentation considérable de l'utilisation des méta-analyses en réseau pour évaluer un large éventail d'interventions de soins de santé (Figure 1.1).

En fait, les médecins et chercheurs ont adopté ces nouvelles méthodes, car elles semblent répondre à leur préoccupation principale, à savoir déterminer le meilleur traitement disponible malgré l'absence d'essais face-face. En outre, les agences nationales d'évaluation des technologies de santé et les organismes de réglementation des médicaments utilisent de plus en plus ces méthodes (Schöttker, Luhmann et al. 2009; Wells, Sultan et al. 2009; Dias, Welton et al. 2011).

Figure 1.1: Nombre de méta-analyses en réseau publiées par an de 2003 à 2012



Données issues d'une revue systématique des méta-analyses en réseau publiées jusqu'au Juillet 2012.

**Les données pour le 2^{ème} semestre 2012 ont été imputées.*

1.4 Biais de publication et méta-analyse en réseau

Les méta-analyses en réseau restent des méta-analyses et devraient être réalisées selon les méthodes rigoureuses utilisées dans les revues systématiques (Li, Puhan et al. 2011). Nous avons

récemment réalisé une revue systématique méthodologique des méta-analyses en réseau publiées (Bafeta, Trinquart et al.). Après interrogation de la Cochrane Database of Systematic Reviews, de la Database of Abstracts of Reviews of Effects, de MEDLINE et EMBASE jusqu'en Juillet 2012, nous avons identifié 121 méta-analyses en réseau. Nos résultats ont montré que les composantes méthodologiques essentielles du processus de revue systématique faisaient souvent défaut dans les rapports de méta-analyses en réseau. En particulier, la possibilité du biais de publication n'était abordée que dans 16 (13%) méta-analyses en réseau.

Alors qu'il existe une littérature statistique abondante consacrée à la méta-analyse en réseau, les biais de publication ont reçu peu d'attention dans ce cadre. En outre, il n'existe pas de test pour la détection du biais de publication ni de méthodes d'analyse de sensibilité au biais de publication pour la méta-analyse en réseau. Ainsi, nos travaux de recherche avaient pour objectif d'évaluer l'impact des biais de publication dans la méta-analyse et de proposer des méthodes de détection et d'analyse de sensibilité des résultats aux biais de publication.

Chapitre 2

Méthodes statistiques de la méta-analyse en réseau

2.1 Comparaisons indirectes ajustées

2.2 Méta-analyse en réseau – modèle à effet fixe

2.3 Méta-analyse en réseau – modèle à effets aléatoires

2.4 Méta-analyse en réseau – paramétrisation de Lu et Ades

2.5 Extension du modèle aux essais à plus de deux bras

2.6 Présentation des résultats d'une méta-analyse en réseau

2.7 Approche bayésienne et approche fréquentiste

2.8 Hypothèses nécessaires

2.9 Evaluation de l'incohérence et modèles d'incohérence

2.1 Comparaisons indirectes ajustées

En l'absence d'essai randomisé entre l'intervention j et l'intervention k , une comparaison indirecte ajustée est possible s'il existe des essais comparant j à b et k à b , où b est un comparateur commun (par exemple, un placebo). Le cas échéant, l'effet de j relativement à k est estimé par $\hat{\theta}_{jk}^{\text{indirecte}} = \hat{\theta}_{jb}^{\text{directe}} - \hat{\theta}_{kb}^{\text{directe}}$ et la variance associée par $\text{var}(\hat{\theta}_{jk}^{\text{indirecte}}) = \text{var}(\hat{\theta}_{jb}^{\text{directe}}) + \text{var}(\hat{\theta}_{kb}^{\text{directe}})$ (Bucher, Guyatt et al. 1997).

S'il existe un ou plusieurs essais comparant l'intervention j et l'intervention k , on peut privilégier l'information directe et ne pas considérer l'information indirecte, en considérant que l'essai face-face est la meilleure source d'information. Néanmoins, ces essais ne sont pas indemnes de biais (défauts méthodologiques, biais liés au financement) et il est parfois préférable de documenter le contraste entre deux traitements sur la base d'une comparaison indirecte ajustée

(Song, Harvey et al. 2008; Alkhafaji, Trinquart et al. 2012). Par ailleurs, parce que les essais face-face font souvent défaut, la quantité d'information indirecte est souvent plus importante que la quantité d'information directe disponible. Si les deux sources d'information sont cohérentes, il est alors intéressant de les combiner, par exemple par une moyenne pondérée par l'inverse de la

$$\text{variance } \hat{\theta}_{jk}^{\text{combiné}} = \left(\frac{\theta_{jk}^{\text{indirecte}}}{\text{var}(\theta_{jk}^{\text{indirecte}})} + \frac{\theta_{jk}^{\text{directe}}}{\text{var}(\theta_{jk}^{\text{directe}})} \right) / \left(\frac{1}{\text{var}(\theta_{jk}^{\text{indirecte}})} + \frac{1}{\text{var}(\theta_{jk}^{\text{directe}})} \right).$$

Cette approche ne peut pas être appliquée à des réseaux d'essais plus grands ou incluant au moins une boucle fermée. Dans la suite, nous décrivons l'extension des modèles pour la méta-analyse conventionnelle de sorte qu'ils puissent être appliqués à tout réseau connecté, quels qu'en soient la taille et la complexité.

2.2 Méta-analyse en réseau – modèle à effet fixe

Nous décrivons les méthodes basées sur les données agrégées disponibles pour chaque essai par opposition aux méthodes basées sur les données individuelles (en anglais, *individual patient data network meta-analysis*). Nous décrivons ici le modèle dans le cas d'un critère de jugement binaire.

Le modèle ci-dessous est basé sur le modèle statistique décrit par Higgins et Whitehead, mais le codage et la notation ont été étendus pour s'appliquer à toute structure de réseau, y compris celle où il n'existe pas d'intervention unique à laquelle toutes les autres interventions ont été comparées (Higgins and Whitehead 1996; Whitehead 2002). Dans un réseau comprenant N essais, nous supposons que chaque essai i a comparé les interventions j et k parmi les T interventions disponibles dans le réseau. Pour chaque essai i , on observe r_{ij} et r_{ik} événements parmi n_{ij} et n_{ik} patients randomisés dans le groupe expérimental j et le groupe témoin k , respectivement. Notons que l'intervention témoin k varie d'un essai à l'autre.

Sous l'hypothèse d'une vraisemblance binomiale, le modèle à effet fixe s'écrit:

$$r_{ij} \sim \text{Bin}(p_{ij}, n_{ij})$$

$$r_{ik} \sim \text{Bin}(p_{ik}, n_{ik})$$

$$\text{logit}(p_{ij}) = \mu_i + \theta_{jk}$$

$$\text{logit}(p_{ik}) = \mu_i$$

avec θ_{jk} l'effet vrai de l'intervention j relativement à k commun à tous les essais comparant j à k (i.e., le log du rapport des cotes d'événement avec j relativement à k).

Le modèle repose sur l'hypothèse de cohérence (en anglais, consistency assumption) entre l'information directe et indirecte: si nous considérons l'intervention b comme l'intervention de référence sur l'ensemble du réseau, les effets des interventions j , k , etc. relativement à l'intervention de référence b , θ_{jb} , θ_{kb} , etc., sont considérés comme paramètres basiques, et les autres contrastes, les paramètres fonctionnels, sont dérivés des équations de cohérence suivantes:

$$\theta_{jk} = \theta_{jb} - \theta_{kb} \text{ pour tout } j \neq k \neq b$$

En général, tout sous-ensemble de contrastes peut être sélectionné pour définir les paramètres basiques tant que les lignes correspondantes sur le graphe du réseau forment un arbre couvrant, c'est-à-dire un sous-graphe où tous les nœuds sont connectés mais sans aucune boucle fermée. Dans ce cas, l'ajout d'un contraste, donc d'une ligne sur le graphe, forme un cycle (une boucle fermée) et représente alors un paramètre fonctionnel (Lu and Ades 2006).

Le choix d'une intervention de référence et l'hypothèse de cohérence permet de réduire le problème d'estimation de $T \cdot (T-1)/2$ contrastes deux à deux entre interventions à l'estimation de $(T-1)$ paramètres basiques, en synthétisant l'information directe et indirecte. Notons que l'on fixe θ_{bb} l'effet de l'intervention de référence b relativement à elle-même égal à 0.

2.3 Méta-analyse en réseau – modèle à effets aléatoires

Avec les mêmes notations que précédemment, notamment en considérant le groupe expérimental j et le groupe témoin k dans l'essai i , le modèle à effets aléatoires s'écrit

$$r_{ij} \sim \text{Bin}(p_{ij}, n_{ij})$$

$$r_{ik} \sim \text{Bin}(p_{ik}, n_{ik})$$

$$\text{logit}(p_{ij}) = \mu_i + \theta_{ijk}$$

$$\text{logit}(p_{ik}) = \mu_i$$

$$\theta_{ijk} \sim N(\theta_{jk}, \tau^2)$$

$\theta_{jk} = \theta_{jb} - \theta_{kb}$ pour tout $j \neq k \neq b$ (hypothèse de cohérence via l'intervention de référence b sur l'ensemble du réseau) avec la contrainte $\theta_{bb} = 0$

avec θ_{ijk} l'effet vrai sous-jacent à chaque essai i entre j and k désormais supposé distribué selon une loi normale centrée sur θ_{jk} l'effet moyen vrai de l'intervention j relativement à k . Dans ce modèle à effets aléatoires, nous faisons l'hypothèse que la variance inter-essais est homogène sur l'ensemble du réseau (i.e., $\tau_{jk}^2 = \tau^2$). Si $\tau^2 = 0$, on retrouve le modèle à effet fixe.

Cette hypothèse peut être relaxée (Lu and Ades 2006; Lu and Ades 2009). Dans ce cas, on suppose $\theta_{ijk} \sim N(\theta_{jk}, \tau_{jk}^2)$ sans restriction sur la structure des variances (modèle à hétéroscédasticité). L'inconvénient potentiel en pratique est l'imprécision sur l'estimation des variances inter-essais, le nombre d'essais disponibles pour chaque comparaison étant souvent limité. Pour optimiser l'estimation, on peut supposer l'interchangeabilité des variances inter-essais (Thorlund, Thabane et al. 2013). Par exemple, on peut considérer un paramètre de dispersion commun τ puis supposer que les écarts types τ_{jk} sont distribués selon une distribution de Student tronquée et décentrée de paramètre de décentrage τ , à $N_{jk}-1$ degrés de liberté, avec N_{jk}

le nombre d'essais pour la comparaison entre j et k : $\tau_{jk} \sim t(0, \tau, N_{jk} - 1)$. Une autre approche consiste à contraindre les variances inter-essais de sorte que l'hypothèse de cohérence soit vérifiée non seulement pour les contrastes entre interventions ($\theta_{jk} = \theta_{jb} - \theta_{kb}$ pour tout $j \neq k \neq b$; équation de cohérence d'ordre 1) mais également pour leurs variances: $\tau_{jk}^2 = \tau_{jb}^2 + \tau_{kb}^2 - 2\rho_{jkb} \tau_{jb} \tau_{kb} \leq \tau_{jb}^2 + \tau_{kb}^2$ avec ρ_{jkb} la corrélation entre deux contrastes entre interventions. Ceci conduit à incorporer au modèle à hétéroscédasticité une inégalité triangulaire d'ordre 2: $|\tau_{jk}^2 - \tau_{kb}^2| \leq \tau_{jb}^2 \leq |\tau_{jb}^2 + \tau_{kb}^2|$ (Lu and Ades 2009).

2.4 Méta-analyse en réseau – paramétrisation de Lu et Ades

Lu et Ades ont décrit une paramétrisation différente basée sur l'extension du modèle hiérarchique de Smith, Spiegelhalter et Thomas pour la méta-analyse conventionnelle (Smith, Spiegelhalter et al. 1995; Lu and Ades 2004). Si nous considérons l'intervention b comme l'intervention de référence sur l'ensemble du réseau, le modèle à effets aléatoires s'écrit ainsi:

$$r_{ij} \sim \text{Bin}(p_{ij}, n_{ij}) \text{ pour } j \neq b$$

$$\text{logit}(p_{ij}) = \mu_i + (T - 1) \cdot \theta_{ijb} / T - \sum_{l \neq j, b} \theta_{ilb} / T \text{ pour } j \neq b$$

$$\text{logit}(p_{ib}) = \mu_i - \sum_{j \neq b} \theta_{ijb} / T \text{ pour l'intervention de référence } b \text{ sur l'ensemble du réseau}$$

$$\theta_{ijb} \sim N(\theta_{jb}, \tau^2) \text{ pour } j \neq b$$

$$\theta_{jk} = \theta_{jb} - \theta_{kb} \text{ pour tout } j \neq k \neq b \text{ (hypothèse de cohérence) avec la contrainte } \theta_{bb} = 0.$$

2.5 Extension du modèle aux essais à plus de deux bras

Nous revenons au modèle statistique décrit par Higgins et Whitehead et présentons une extension du modèle pour prendre en compte les essais randomisés à plus de deux bras. Il est

fréquent qu'un réseau d'essais inclue au moins un essai à plus de deux bras. Sur un échantillon représentatif de 519 essais randomisés publiés en Décembre 2000 et indexés dans Medline, 27% avaient plus de deux bras de randomisation (Chan, Altman 2005 Lancet).

Dans un essai i à trois bras comparant les interventions A , B et C , trois contrastes entre interventions sont estimables. Si ces contrastes sont écrits sous forme de vecteurs, deux d'entre eux, mais pas les trois, seraient linéairement indépendants; autrement dit, chacun d'eux peut être écrit en fonction des deux autres. Par conséquent, seulement deux des trois contrastes doivent être inclus dans un modèle. Sans perte de généralité, nous supposons que l'intervention témoin est C , et que l'essai i fournit les estimations de θ_{iAC} et θ_{iBC} . Toutefois, étant donné que ces deux contrastes ont en commun le traitement C , les contrastes estimés ne sont pas statistiquement indépendants.

La distribution conjointe de θ_{iAC} et θ_{iBC} peut s'écrire comme une distribution normale bivariée:

$$\begin{pmatrix} \theta_{iAC} \\ \theta_{iBC} \end{pmatrix} \sim N \left(\begin{pmatrix} \theta_{AC} \\ \theta_{BC} \end{pmatrix}, \begin{pmatrix} \tau_{AC}^2 & \delta \\ \delta & \tau_{BC}^2 \end{pmatrix} \right) \text{ avec } \delta = \text{cov}(\theta_{iAC}, \theta_{iBC})$$

Nous faisons l'hypothèse d'une variance homogène pour chaque comparaison deux à deux, i.e., $\tau_{AC}^2 = \tau_{BC}^2 = \tau_{AB}^2 = \tau^2$. Les distributions marginales des contrastes s'écrivent $\theta_{iAB} \sim N(\theta_{AB}, \tau^2)$, $\theta_{iAC} \sim N(\theta_{AC}, \tau^2)$ et $\theta_{iBC} \sim N(\theta_{BC}, \tau^2)$. Dans cet essai à trois bras, l'hypothèse de cohérence est inhérente, par exemple $\theta_{iAB} = \theta_{iAC} - \theta_{iBC}$, et donc $\text{var}(\theta_{iAB}) = \text{var}(\theta_{iAC}) + \text{var}(\theta_{iBC}) - 2 \text{cov}(\theta_{iAC}, \theta_{iBC})$. Ceci implique alors que la covariance entre deux contrastes quelconques est égale à $\tau^2/2$. Par conséquent, on peut décrire la distribution des contrastes dans l'essai i par une loi normale bivariée:

$$\begin{pmatrix} \theta_{iAC} \\ \theta_{iBC} \end{pmatrix} \sim N \left(\begin{pmatrix} \theta_{AC} \\ \theta_{BC} \end{pmatrix}, \begin{pmatrix} \tau^2 & \tau^2/2 \\ \tau^2/2 & \tau^2 \end{pmatrix} \right)$$

et la distribution de θ_{iAB} est déduite de l'équation de cohérence $\theta_{iAB} = \theta_{iAC} - \theta_{iBC}$.

On peut réécrire cette distribution normale bivariée comme deux distributions normales univariées inter-reliées, i.e., une distribution normale marginale pour θ_{iAC} et une distribution normale conditionnelle à θ_{iAC} pour θ_{iBC} (Spiegelhalter 1998):

$$\theta_{iAC} \sim N(\theta_{AC}, \tau^2)$$

$$\theta_{iBC} | \theta_{iAC} \sim N(\theta_{BC} + (\theta_{iAC} - \theta_{AC})/2, 3 \cdot \tau^2/4)$$

Cette approche peut être généralisée à tout type d'essai, quel que soit le nombre de bras de randomisation. En effet, dans un réseau évaluant T interventions, on peut considérer tout essai du réseau comme un essai à T bras, les quantités d'effet fournies par l'essai sont des réalisations d'une distribution normale multivariée, et les traitements qui ne sont pas rapportés sont manquants au hasard (en anglais, missing at random). Notons que si les traitements sont manquants au hasard sans tenir compte de l'efficacité du traitement, cela ne signifie pas que les différents traitements peuvent apparaître dans un essai de manière équiprobable.

Pour un réseau de N essais évaluant T interventions, on note M le nombre maximum de bras de randomisation par essai dans le réseau. Les M vecteurs $t_{i,k}$, $k=1, \dots, M$, de longueurs N , décrivent quelle intervention correspond au groupe k de chaque essai i . Par convention, $t_{i,1}$ décrit l'intervention témoin de l'essai i . Les M vecteurs $r_{i,k}$ et $n_{i,k}$, $k=1, \dots, M$, de longueurs N , décrivent les nombres d'événements et de patients randomisés dans le groupe k de chaque essai i . Le modèle dans ce cas général s'écrit:

$$r_{i,k} \sim \text{Bin}(p_{i,t_{i,k}}, n_{i,k})$$

$$\text{logit}(p_{i,t_{i,1}}) = \mu_i \text{ pour le groupe témoin de l'essai } i$$

$$\text{logit}(p_{i,t_{i,k}}) = \mu_i + \theta_{i,t_{i,k}t_{i,1}} \text{ pour } k \neq 1$$

$$\begin{pmatrix} \theta_{i,t_{i,2}t_{i,1}} \\ \vdots \\ \theta_{i,t_{i,M}t_{i,1}} \end{pmatrix} \sim MVN \left(\begin{pmatrix} \theta_{t_{i,2}t_{i,1}} \\ \vdots \\ \theta_{t_{i,M}t_{i,1}} \end{pmatrix}, \begin{pmatrix} \tau^2 & \tau^2/2 & \dots & \tau^2/2 \\ \tau^2/2 & \tau^2 & \dots & \tau^2/2 \\ \vdots & \vdots & \ddots & \vdots \\ \tau^2/2 & \tau^2/2 & \dots & \tau^2 \end{pmatrix} \right)$$

$\theta_{t_{i,k}t_{i,1}} = \theta_{t_{i,k}b} - \theta_{t_{i,1}b}$ pour tout $t_{i,1}, t_{i,k} \neq b$ (hypothèse de cohérence via l'intervention de référence b sur l'ensemble du réseau) avec la contrainte $\theta_{bb} = 0$.

Comme précédemment, on peut écrire la distribution normale marginale de $\theta_{i,t_{i,2}t_{i,1}}$ et la distribution normale conditionnelle de $\theta_{i,t_{i,k}t_{i,1}}$ pour $k > 2$:

$$\theta_{i,t_{i,2}t_{i,1}} \sim N(\theta_{t_{i,2}t_{i,1}}, \tau^2)$$

$$\theta_{i,t_{i,k}t_{i,1}} \mid \begin{pmatrix} \theta_{i,t_{i,2}t_{i,1}} \\ \vdots \\ \theta_{i,t_{i,k-1}t_{i,1}} \end{pmatrix} \sim N \left(\theta_{t_{i,k}t_{i,1}} + \frac{1}{k-1} \sum_{j=2}^{k-1} (\theta_{i,t_{i,j}t_{i,1}} - \theta_{t_{i,j}t_{i,1}}), \frac{k}{2(k-1)} \tau^2 \right)$$

2.6 Présentation des résultats d'une méta-analyse en réseau

Les résultats de la méta-analyse en réseau s'expriment sous la forme de quantités d'effet: pour T interventions, on estime les $T \times (T-1)/2$ quantités d'effet (rapports des cotes dans le cas d'un critère de jugement binaire) pour toutes les comparaisons deux à deux entre les interventions évaluées. Chaque estimé est accompagné d'un intervalle de confiance (ou de crédibilité) autour de la quantité d'effet vraie. Ces résultats peuvent être rapportés sous forme d'une matrice $T \times T$ ou représentés graphiquement sur un forest plot.

En particulier, il est utile de considérer les comparaisons deux à deux pour lesquelles il existe de l'information directe (i.e., au moins un essai comparant les deux interventions) et de comparer les quantités d'effet estimées via la méta-analyse en réseau à celles obtenues via la méta-analyse conventionnelle. En examinant chaque estimé et l'intervalle de confiance associé, on peut ainsi évaluer la manière dont l'information indirecte influence ou non l'information

directe. Par ailleurs, il est important de rapporter une mesure d'hétérogénéité, en particulier la variance inter-essais τ^2 .

D'autre part, la direction et l'intensité de l'effet d'une intervention détermine de façon significative l'ordre de préférence parmi toutes les interventions évaluées. Il est possible d'ordonner les interventions de la plus efficace à la moins efficace en utilisant des probabilités. Ainsi, la probabilité que l'intervention k soit la t -ième meilleure intervention parmi les T interventions évaluées est définie comme $P(k = t)$, $t=1,2, \dots, T$. Nous appelons ces probabilités les probabilités de rang. Dans l'approche bayésienne, à chaque itération dans la méthode de Monte Carlo par chaînes de Markov, chaque traitement peut être classé en fonction de la quantité d'effet estimée relativement à l'intervention de référence b du réseau :

$$\text{Rang}(k) = \begin{cases} \sum_{t=1}^T I(\theta_{tb} \leq \theta_{kb}) & \text{si le critère est négatif} \\ T + 1 - \sum_{t=1}^T I(\theta_{tb} \leq \theta_{kb}) & \text{si le critère est positif} \end{cases}$$

Ensuite, la proportion des itérations pour lesquelles une intervention donnée se classe première parmi toutes les interventions donne la probabilité d'être au premier rang pour ce traitement. Des probabilités similaires sont estimées pour être au deuxième rang, au troisième, et ainsi de suite. Dans l'approche fréquentiste, ces probabilités peuvent être estimées via la méthode de ré-échantillonnage bootstrap.

Pour chaque intervention et chaque rang, la somme de ces probabilités vaut 1. On peut aussi estimer les probabilités de rang cumulées, qui sont les probabilités qu'une intervention soit parmi les meilleurs traitements. On peut représenter graphiquement la hiérarchie des traitements en utilisant une courbe des probabilités de rang (en anglais, rankograms) ou une courbe des probabilités de rang cumulées pour chaque traitement.

D'autres indicateurs peuvent être dérivés de ces probabilités, comme la surface sous la courbe des probabilités de rang cumulées pour chaque traitement (en anglais, surface under the

cumulative ranking SUCRA) : cette aire vaut 1 lorsqu'il est certain que l'intervention considérée est la meilleure et 0 lorsqu'il est certain qu'elle est la pire (Salanti, Ades et al. 2010). Enfin, on peut estimer le rang médian de chaque intervention et l'intervalle de confiance (crédibilité) associé au vrai rang.

2.7 Approche bayésienne et approche fréquentiste

L'approche bayésienne a été la première décrite et est la plus fréquemment utilisée (Higgins and Whitehead 1996; Whitehead 2002; Lu and Ades 2004). Dans le cadre des méta-analyses et des méta-analyses en réseau, elle repose sur l'échantillonnage de Gibbs, une classe de méthodes Monte-Carlo par chaînes de Markov (MCMC). C'est l'approche que nous avons adoptée dans la suite. Dans ce cadre, on définit des lois a priori non informatives pour les paramètres basiques θ_{jb} et pour la variance inter-essais τ^2 supposée homogène: $\theta_{jb} \sim N(0, 100^2)$ et $1/\tau^2 \sim \text{Gamma}(0.001, 0.001)$ ou $\tau \sim \text{Uniforme}(0, 2)$ par exemple si l'effet du traitement est mesuré par un log rapport de cote (Lambert, Sutton et al. 2005). Le choix de lois a priori non informatives implique que le choix du traitement de référence aura un impact négligeable sur les lois a posteriori; de plus, les résultats de l'approche bayésienne seront similaires à ceux de l'approche fréquentiste. Cette dernière est techniquement possible mais rarement utilisée. Plusieurs techniques ont été décrites, notamment l'ajustement de modèles linéaires généralisés ou de modèles de méta-régression (Lumley 2002; Salanti, Higgins et al. 2008; Stijnen, Hamza et al. 2010; Jones, Roger et al. 2011; Senn, Gavini et al. 2012; White, Barrett et al. 2012).

2.8 Hypothèses nécessaires

Dans un modèle à effets aléatoires pour une méta-analyse conventionnelle, nous admettons en fait l'hypothèse d'interchangeabilité, c'est-à-dire que l'intensité des effets du traitement d'un essai à l'autre ne peut être différenciée a priori. Formellement, la distribution conjointe des effets sous-jacents $p(\theta_1, \dots, \theta_n)$ est identique pour toute permutation Π des indices des essais: $p(\theta_{\Pi(1)}, \dots, \theta_{\Pi(n)})$. Si les effets vrais sous-jacents sont supposés interchangeables, il s'ensuit que les θ_i sont indépendants et identiquement distribués: dans une méta-analyse conventionnelle de n_{jb} essais, l'effet traitement vrai de j relativement à b dans l'essai i est distribué selon une loi commune $\theta_{ijb} \sim N(\theta_{jb}, \tau^2)$ et, dans une autre méta-analyse conventionnelle de n_{kb} essais, l'effet traitement vrai de k relativement à b dans l'essai i est distribué selon une loi commune $\theta_{ikb} \sim N(\theta_{kb}, \tau^2)$ (Higgins, Thompson et al. 2009).

La seule hypothèse supplémentaire dans une méta-analyse en réseau par rapport à la méta-analyse conventionnelle est que ces hypothèses d'interchangeabilité s'étendent à l'ensemble des $n_{jb}+n_{kb}$ essais. Si maintenant n_{jk} essais comparant j à k sont ajoutés au réseau, et que les deux hypothèses précédentes sont étendues aux n_{jk} essais, alors la cohérence requise entre les contrastes jb , kb et jk est garantie automatiquement (Lu and Ades 2009; Ades 2011). Autrement dit, l'interchangeabilité suppose que si tous les essais avaient inclus tous les traitements évalués dans le réseau, chaque essai conduirait à estimer les mêmes quantités d'effet, échangeables, entre traitements.

D'autres ont formalisé les hypothèses requises sous la forme de la cohérence, l'homogénéité et la similarité (Song, Loke et al. 2009; Hoaglin, Hawkins et al. 2011; Jansen, Fleurence et al. 2011). L'hypothèse de cohérence - qui découle de l'hypothèse d'interchangeabilité - désigne, pour l'estimation de la quantité d'effet entre deux traitements donnés, l'absence de

discordance entre les résultats issus de comparaisons indirectes et ceux issus de comparaisons directes. Plusieurs revues systématiques des comparaisons indirectes ajustées et méta-analyses en réseau existantes ont montré que l'incohérence (en anglais, inconsistency) entre données indirectes et directes était peu fréquente (Song, Altman et al. 2003; Glenny, Altman et al. 2005; Song, Xiong et al. 2011). Néanmoins, ces résultats empiriques restent limités et des outils statistiques ont été proposés pour évaluer l'incohérence et ajuster les modèles sur l'incohérence. L'hypothèse d'homogénéité - déjà nécessaire dans la méta-analyse classique - désigne l'absence de variabilité excessive dans les effets traitement entre les différents essais du réseau. Enfin, l'hypothèse de similarité désigne l'absence de différence excessive entre les comparateurs communs dans l'évaluation de l'effet relatif par comparaison indirecte (Baker and Kramer 2002; Xiong, Parekh-Bhurke et al. 2013). L'hétérogénéité est évaluable sur chaque branche du réseau, l'incohérence est partiellement évaluable (sur chaque boucle fermée) mais la similarité est difficilement évaluable (Xiong, Parekh-Bhurke et al. 2013).

2.9 Evaluation de l'incohérence et modèles d'incohérence

2.9.1 Méthode de Bucher

Pour toute boucle fermée jkb du réseau, on peut estimer $\hat{\theta}_{jk}^{\text{directe}}$ par synthèse de l'information directe et $\hat{\theta}_{jk}^{\text{indirecte}} = \hat{\theta}_{jb}^{\text{directe}} - \hat{\theta}_{kb}^{\text{directe}}$ par comparaison indirecte ajustée. On peut ensuite estimer l'incohérence sur la boucle par $\omega_{jkb} = \hat{\theta}_{jk}^{\text{directe}} - \hat{\theta}_{jk}^{\text{indirecte}}$. La variance associée est $\text{var}(\omega_{jkb}) = \text{var}(\hat{\theta}_{jk}^{\text{directe}}) + \text{var}(\hat{\theta}_{jk}^{\text{indirecte}})$. On peut alors tester l'hypothèse nulle de cohérence sur la boucle en utilisant la statistique $z_{jkb} = \omega_{jkb} / \text{var}(\omega_{jkb})$.

Dans des réseaux avec plusieurs boucles, cette approche peut être appliquée de manière répétée à toutes les boucles fermées (ou toutes les boucles fermées indépendantes) du réseau.

Cependant, le test est peu puissant et l'approche soulève un problème de multiplicité des tests (Dias, Welton et al. 2010; Dias, Welton et al. 2012).

Une autre possibilité consiste à ajuster des modèles d'incohérence et de comparer la qualité d'ajustement du modèle d'incohérence à celui d'un modèle cohérent. Nous décrivons ci-après le modèle d'incohérence de Lumley et celui de Lu et Ades.

2.9.2 *Modèle de Lumley*

Dans un réseau comprenant N essais, nous supposons que chaque essai i a comparé les interventions j et k parmi les T interventions disponibles dans le réseau. Chaque essai a fourni une estimation de l'effet y_{ijk} de l'intervention expérimentale j relativement à l'intervention témoin k et la variance associée v_{ijk} . Comme précédemment, sous l'hypothèse d'une vraisemblance normale, $y_{ijk} \sim N(\theta_{ijk}, v_{ijk})$ où θ_{ijk} est la quantité d'effet vraie sous-jacente à chaque essai comparant les interventions j et k .

Dans le modèle de Lumley, on décompose θ_{ijk} de la manière suivante (Lumley 2002):

$$\theta_{ijk} = \theta_{jk} + \eta_{ijk} + \omega_{jk}$$

$$\eta_{ijk} \sim N(0, \tau^2)$$

$$\omega_{jk} \sim N(0, \sigma_\omega^2)$$

de sorte que $\theta_{jk} = \mu_j - \mu_k$ est l'effet moyen de l'intervention j relativement à k , η_i est un terme d'interaction essai-traitement qui reflète l'hétérogénéité entre essais comparant j et k (i.e., la variabilité vraie τ^2 dans les effets entre j et k). Enfin, ω_{jk} capture la modification de l'effet absolu de l'intervention j lorsqu'elle est comparée spécifiquement à l'intervention k . Afin de synthétiser différentes comparaisons entre interventions, il est nécessaire que l'effet absolu μ_j de l'intervention j soit le même qu'elle que soit l'intervention à laquelle j est comparée; autrement

dit, ω_{jk} doit être proche de zéro. Ainsi ω_{jk} est un facteur d'incohérence spécifique à chaque comparaison deux à deux entre traitements (plutôt qu'à chaque essai) et capture l'incohérence de cette paire d'intervention avec le reste de l'information apportée par le réseau. Ceci permet aux comparaisons indirectes d'être incohérentes par rapport aux comparaisons directes d'une même paire d'interventions. On suppose que l'ensemble des facteurs d'incohérence sont distribués selon une loi normale. Lumley mesure l'incohérence du réseau via l'écart type σ_ω des facteurs d'incohérence. Notons que le modèle repose sur l'hypothèse d'homogénéité des variances (i.e., $\tau_{jk}^2 = \tau^2$ et $\sigma_{\omega_{jk}}^2 = \sigma_\omega^2$). De surcroît, Lumley modélise l'hétérogénéité en spécifiant $\text{var}(y_{ijk} | \eta, \omega) = a(b + \sqrt{v_{ijk}})^2$. Finalement, on spécifie la contrainte $\mu_b = 0$ pour l'intervention de référence du réseau.

2.9.3 Modèle de Lu et Ades

A la manière de Lumley, Lu et Ades ont décrit un modèle autorisant l'incohérence entre information directe et indirecte (Lu and Ades 2006). Le modèle est basé sur le modèle de Higgins et Whitehead étendu, décrit précédemment. Dans un réseau comprenant N essais, nous supposons que chaque essai i a comparé les interventions j et k parmi les T interventions disponibles dans le réseau.

Pour décrire l'incohérence, Lu et Ades ont introduit des facteurs d'incohérence w reliant les paramètres fonctionnels aux paramètres basiques: $\theta_{jk} = \theta_{jb} - \theta_{kb} + \omega_{bjk}$ pour $j, k \neq b$. On suppose de plus que ces facteurs d'incohérence sont distribués selon une distribution normale commune: $\omega_{bjk} \sim N(0, \sigma_\omega^2)$ pour $j, k \neq b$

Il apparaît clairement que ces facteurs d'incohérence sont liés à des boucles et non à des arêtes du graphe du réseau. Par conséquent, un facteur d'incohérence ω_{bjk} reflètera globalement

le degré d'incohérence entre les sources d'information sur la boucle $bjkb$, i.e., entre les quantités d'effet estimées par les essais j versus b , k versus b et j versus k , et non pas le biais dans l'estimation de θ_{jk} . Par ailleurs, cette incohérence ne peut apparaître, et donc ne peut être évaluée, que s'il existe trois ensembles d'essais pour chaque comparaison; autrement dit, s'il existe de l'information directe et indirecte pour la comparaison j versus k .

Le nombre d'incohérences potentielles dans le réseau, i.e. le nombre de facteurs d'incohérence, correspond au nombre de boucles fermées indépendantes dans le réseau. Si dans notre réseau comprenant N essais évaluant T interventions, on compte K comparaisons parmi $T \cdot (T-1)/2$ pour lesquelles on dispose d'information directe, le nombre d'incohérences potentielles vaut $K - T + 1$. Lu et Ades ont appelé ce nombre de boucles fermées indépendantes le nombre de degrés de liberté d'incohérence (en anglais, inconsistency degrees of freedom).

Le modèle incohérent s'écrit donc

$$r_{ij} \sim \text{Bin}(p_{ij}, n_{ij})$$

$$r_{ik} \sim \text{Bin}(p_{ik}, n_{ik})$$

$$\text{logit}(p_{ij}) = \mu_i + \theta_{ijk}$$

$$\text{logit}(p_{ik}) = \mu_i$$

$$\theta_{ijk} \sim N(\theta_{jk}, \tau^2)$$

$$\theta_{jk} = \theta_{jb} - \theta_{kb} + \omega_{bjk} \text{ pour } j, k \neq b \text{ avec la contrainte } \theta_{bb} = 0$$

$$\omega_{bjk} \sim N(0, \sigma_\omega^2)$$

Pour quantifier l'incohérence globale du réseau, on peut, à la manière de Lumley considérer la variance σ_ω^2 des facteurs d'incohérence et estimer la probabilité que σ_ω^2 soit supérieure à l'hétérogénéité inter-essais τ^2 .

2.9.4 Essais à plus de deux bras et incohérence

La présence d'essais à plus de deux bras dans un réseau rend la définition et la modélisation de l'incohérence plus complexe. En effet, l'incohérence liée à une boucle comme définie précédemment ne peut pas survenir dans un essai à plus de deux bras. Il faut donc distinguer l'incohérence liée à une boucle et l'incohérence liée au plan d'expérience (en anglais, design). Cette dernière correspond à une différence entre les contrastes estimés via l'essai multi-bras et au moins un des contrastes estimés via les essais à deux bras sur la même boucle. L'estimation du contraste θ_{jk} va résulter de la synthèse d'informations potentiellement incohérentes (par exemple, ensembles d'essais à deux bras sur la boucle $bjkb$) et d'informations cohérentes structurellement (par exemple, essais à trois bras sur la même boucle $bjkb$) mais potentiellement incohérentes avec les précédentes.

Le modèle de Lumley a été développé pour des réseaux d'essais à deux bras et il n'a pas été étendu aux réseaux incluant des essais multi-bras. Le modèle de Lu et Ades permet de tenir compte de ces essais. L'inconvénient est qu'il n'existe pas d'algorithme général pour paramétrer le modèle et que l'écriture du modèle n'est pas unique mais dépend du choix des paramètres basiques et de l'ordre des traitements. Higgins et al ont récemment décrit un modèle d'interaction traitement-plan d'expérience qui permet de modéliser l'incohérence dans un cadre plus général (Higgins, Jackson et al. 2012).

Chapitre 3

Impact du biais de publication dans la méta-analyse en réseau

3.1 Contexte

3.2 Méthodes

3.3 Résultats

3.4 Discussion

3.5 Conclusion

3.1 Contexte

Le programme de « Comparative Effectiveness Research » (CER) est apparu comme ayant un fort potentiel pour permettre des améliorations en matière de santé. La CER est définie comme « la génération et la synthèse de données probantes qui compare les avantages et les inconvénients des différentes méthodes pour prévenir, diagnostiquer, traiter et surveiller une maladie donnée » (Iom 2009; Dickersin 2010). Souvent, les nombreuses approches thérapeutiques disponibles pour une maladie donnée n'ont jamais été comparées face-à-face par des essais contrôlés randomisés (Volpp and Das 2009; Hochman and McCormick 2010; Lathyris, Patsopoulos et al. 2010; Estellat and Ravaud 2012). Contrairement aux méta-analyses conventionnelles, qui permettent d'évaluer si une intervention spécifique est efficace (relativement à une intervention témoin donnée), les méta-analyses en réseau permettraient de mieux répondre à la question posée par l'ensemble des professionnels de la santé: quelle est la meilleure intervention parmi les différentes interventions existantes pour une condition particulière?

Dans ce cadre, l'intervention A est comparée à un comparateur commun C, puis l'intervention B à C, et la comparaison indirecte ajustée est alors présumée permettre la comparaison de A à B malgré l'absence de tout essai face-face entre les deux interventions. Une méta-analyse en réseau permet l'analyse simultanée de plusieurs interventions concurrentes via la synthèse de comparaisons directes et indirectes (Lumley 2002; Lu and Ades 2004). L'avantage réside dans la possibilité d'estimer les quantités d'effet pour toutes les comparaisons deux à deux possibles entre interventions et d'établir un classement de ces interventions. Les dernières années ont vu une augmentation considérable de l'utilisation des méta-analyses en réseau pour évaluer un large éventail d'interventions de soins de santé (Glenny, Altman et al. 2005; Song, Loke et al. 2009). Ces méthodes pourraient être particulièrement utiles dans le cadre de la CER (Kent and Thaler 2008; Stafford, Wagner et al. 2009) mais, avant leur diffusion plus large, une évaluation approfondie de leurs limites est nécessaire.

Le biais de publication est une des principales menaces à la validité des résultats des revues systématiques ou méta-analyses conventionnelles (Montori, Ioannidis et al. 2008; Hopewell, Loudon et al. 2009; Eyding, Lelgemann et al. 2010; Godlee and Loder 2010; Song, Parekh et al. 2010). Le biais de publication désigne spécifiquement le fait que les résultats d'une étude dans leur intégralité ne soient pas publiés. En fait, ce terme englobe désormais d'autres types de biais, notamment la présentation sélective des analyses statistiques, lorsque les résultats de certaines analyses statistiques sont présentés de manière sélective, à la fois en fonction de l'intensité et de la direction des résultats (Montori, Ioannidis et al. 2008). Plusieurs études ont montré que la base de données de la Food and Drug Administration (FDA) offrait des possibilités intéressantes pour l'étude des biais de publication (Turner, Matthews et al. 2008; Vedula, Bero et al. 2009; Hart, Lundh et al. 2012). Ces biais ont reçu peu d'attention dans le cadre de la méta-

analyse en réseau. Nous avons donc cherché à évaluer l'impact des biais de publication sur les résultats de la méta-analyse en réseau.

3.2 Méthodes

Nous avons utilisé des jeux de données créés à partir des rapports d'évaluation par la FDA d'essais d'antidépresseurs et à partir de leurs publications correspondantes. Pour chaque jeu de données, nous avons utilisé une méta-analyse en réseau pour estimer toutes les comparaisons deux à deux possibles entre ces médicaments. Ces jeux de données différaient parce que les résultats de certains essais n'ont pas été publiés, en fonction de la nature des résultats. En outre, dans certains articles publiés, des analyses spécifiques ont été rapportées de manière sélective et les quantités d'effets différaient de celle de la FDA. En comparant les résultats de la méta-analyse en réseau des essais publiés à ceux de la méta-analyse en réseau des essais enregistrés par la FDA, nous avons évalué l'impact des biais de publication globalement. Pour évaluer l'impact du biais de publication spécifiquement, nous avons comparé les résultats de la méta-analyse en réseau des essais publiés mais avec les quantités d'effet de la FDA à ceux de la méta-analyse en réseau des essais enregistrés par la FDA. Pour évaluer l'impact de la présentation sélective des analyses statistiques, nous avons comparé les résultats de la méta-analyse en réseau des essais publiés (avec les quantités d'effet publiées) à ceux de la méta-analyse en réseau des essais publiés mais avec les quantités d'effet de la FDA.

3.2.1 Données issues de la FDA et des articles publiés

Les jeux de données que nous avons utilisés ont été décrits et publiés précédemment par Turner et al. (Tableau C de l'annexe) (Turner, Matthews et al. 2008). En bref, les auteurs ont

identifié tous les essais contrôlés randomisés versus placebo de 12 antidépresseurs approuvés par la FDA. Puis ils ont identifié les publications correspondant à ces essais en cherchant les bases de données de la littérature et en contactant les promoteurs de ces essais. Dans la base de données de la FDA, les auteurs ont identifié 74 essais comparant un des 12 antidépresseurs à placebo (portant sur 12 564 patients). Parmi les 74 essais, les résultats de 23 essais (portant sur 2903 patients) n'ont pas été publiés. Les auteurs ont extrait les quantités d'effet estimées rapportées dans les articles pour les 51 essais avec résultats publiés et les quantités d'effet estimées rapportées dans les rapports d'évaluation de la FDA pour les 74 essais enregistrés par la FDA. Les données ont été extraites des articles et des revues FDA de manière indépendante et en double, avec résolution des désaccords par un processus consensus.

Le critère de jugement était la modification entre la randomisation et le suivi du score évaluant la sévérité des épisodes de dépression. Parce que la dépression a été évaluée par l'échelle de dépression de Hamilton ou l'échelle de Montgomery-Asberg, la mesure d'effet traitement était une différence moyenne standardisée (i.e., la différence des modifications moyennes entre les groupes antidépresseur et placebo divisée par un écart type commun).

3.2.2 Méta-analyses en réseau

Nous avons effectué des méta-analyses en réseau en utilisant un modèle hiérarchique à effets aléatoires avec une approche bayésienne (Higgins and Whitehead 1996; Lu and Ades 2004; Salanti, Marinho et al. 2009; Welton, Caldwell et al. 2009). Le modèle a permis d'estimer les quantités d'effet pour les 66 comparaisons deux à deux possibles entre les 12 agents antidépresseurs ($12 \times 11/2$, à savoir, 66 différences moyennes standardisées pour toutes les paires d'agents). Pour chaque comparaison deux à deux entre médicaments, nous avons estimé la quantité d'effet et l'intervalle de crédibilité associé par la médiane et les percentiles à 5% et 95%

de la loi a posteriori, respectivement. Le modèle est décrit en détail en Annexe 1, Texte 1. Nous avons également calculé la probabilité que chaque agent antidépresseur soit le meilleur (Salanti, Ades et al. 2010). Nous avons établi un classement des médicaments concurrents par la médiane de la distribution a posteriori pour le rang de chaque médicament. Pour toutes les analyses, nous avons arbitrairement orienté chaque comparaison (A versus B ou B versus A) de sorte que la quantité d'effet estimée par la méta-analyse en réseau des 74 essais enregistrés par la FDA soit positive.

Nous avons considéré que la significativité statistique au seuil de 0.05 était atteinte lorsque l'intervalle de crédibilité à 95% ne comprenait pas 0. Nous avons utilisé WinBUGS v1.4.3 (Imperial College, Londres, Royaume-Uni) pour estimer tous les modèles bayésiens, et R v2.11.1 (R Development Core Team, Vienne, Autriche) pour résumer les résultats et évaluer la convergence. Le code WinBUGS est décrit en détail en Annexe 1, Texte 2.

3.2.3 Impact des biais de publication

Pour évaluer l'impact des biais de publication, nous avons comparé les résultats de la méta-analyse en réseau des 74 essais enregistrés par la FDA, avec les quantités d'effet extraites des rapport d'évaluation de la FDA, considérés comme les estimations de référence, à ceux de la méta-analyse en réseau des 51 essais avec résultats publiés, avec les quantités d'effet extraites des articles publiés.

Premièrement, nous avons dessiné le nuage de points des 66 quantités d'effet deux à deux dérivées d'une méta-analyse en réseau par rapport à l'autre. A la manière de Cohen, nous avons examiné si l'intensité des quantités d'effet deux à deux était faible ($<0,2$) ou modérée ($0,2-0,5$) pour évaluer approximativement la significativité clinique (Cohen 1988). Deuxièmement, nous avons calculé les 66 différences relatives entre quantités d'effet deux à deux dérivées d'une méta-

analyse en réseau par rapport à l'autre: $|(SMD^{publi\acute{e}} - SMD^{FDA})/SMD^{FDA}|$. Troisièmement, nous avons résumé le nombre de fois où les quantités d'effet deux à deux dérivées de la méta-analyse en réseau des 51 essais publiés et des 74 essais enregistrés par la FDA différaient en valeur absolue d'au moins 100% et 50%, respectivement. Une différence en valeur absolue d'au moins 50% signifie que la quantité d'effet issue des données publiées est moins de la moitié, ou plus d'une fois et demie, la quantité d'effet issue des données FDA, ce qui indique des différences importantes dans l'estimation. Une différence en valeur absolue d'au moins 100% signifie que la quantité d'effet issue des données publiées est négative (alors que la quantité d'effet issue des données FDA est positive) ou est plus de deux fois la quantité d'effet issue des données FDA, ce qui indique des différences considérables dans l'estimation. Enfin, nous avons comparé les probabilités que chaque agent antidépresseur soit le meilleur et le classement des médicaments obtenus par chaque méta-analyse en réseau.

3.2.4 Impact du biais de publication affectant un seul médicament

Nous avons évalué de manière hypothétique la façon dont le biais de publication affectant un seul médicament peut influencer sur le classement de tous les médicaments. Nous avons effectué successivement 12 méta-analyses en réseau, en supposant que le biais de publication n'affectait qu'un seul médicament, chacun à son tour. Pour le médicament supposé être affecté par le biais de publication, nous avons utilisé les essais publiés et leurs quantités d'effet publiées, alors que pour les 11 autres médicaments, nous avons utilisé les essais enregistrés par la FDA et les quantités d'effet issues des rapports d'évaluation de la FDA. Nous avons ensuite comparé les probabilités que chaque agent antidépresseur soit le meilleur et le classement des médicaments obtenus à partir de chacune de ces 12 méta-analyses en réseau à ceux obtenus à partir de la méta-analyse en réseau des 74 essais enregistrés par la FDA.

3.2.5 Impact spécifique du biais de publication et de la présentation sélective des analyses

Dans une analyse exploratoire, nous avons cherché à séparer l'impact des différents types de biais de publication. En particulier, la présentation sélective des analyses statistiques peut avoir une influence importante, et il faut convertir relativement peu d'essais négatifs en des essais positifs pour atteindre un biais similaire à celui observé si 10 fois plus d'essais négatifs restaient non publiés (Ioannidis 2011). Ici, l'analyse statistique rapportée dans les articles publiés pouvait différer de celle des rapports d'évaluation de la FDA, qui suivait les méthodes pré-spécifiées (les évaluateurs de la FDA revoient le protocole original soumis avant la conduite de l'essai et des statisticiens ré-analysent les données brutes fournies par le promoteur) (Turner 2005). Ces différences pouvaient résulter d'écarts par rapport à l'analyse en intention de traiter, des méthodes variées pour gérer les données manquantes, de l'analyse d'essais distincts comme s'il s'agissait d'un seul, de la présentation de données provenant de centres uniques dans des essais pourtant multicentriques (Turner, Matthews et al. 2008).

Nous avons évalué l'impact du biais de publication spécifiquement en comparant les résultats de la méta-analyse en réseau des 51 essais publiés mais avec les quantités d'effet de la FDA à ceux de la méta-analyse en réseau des 74 essais inscrits par la FDA. Nous avons supposé que les différences observées seraient attribuables au biais de publication seulement (par construction, la présentation sélective des analyses n'existe plus). Ensuite, nous avons évalué l'impact du biais de présentation sélective des analyses en comparant les résultats de la méta-analyse en réseau des 51 essais publiés (avec les quantités d'effet publiées) à ceux de la méta-analyse en réseau des 51 essais publiés mais avec les quantités d'effet extraites des rapports d'évaluation de la FDA. Nous avons supposé que les différences observées seraient attribuables

au biais de présentation sélective des analyses uniquement (par construction, le biais de publication n'existe plus).

3.3 Résultats

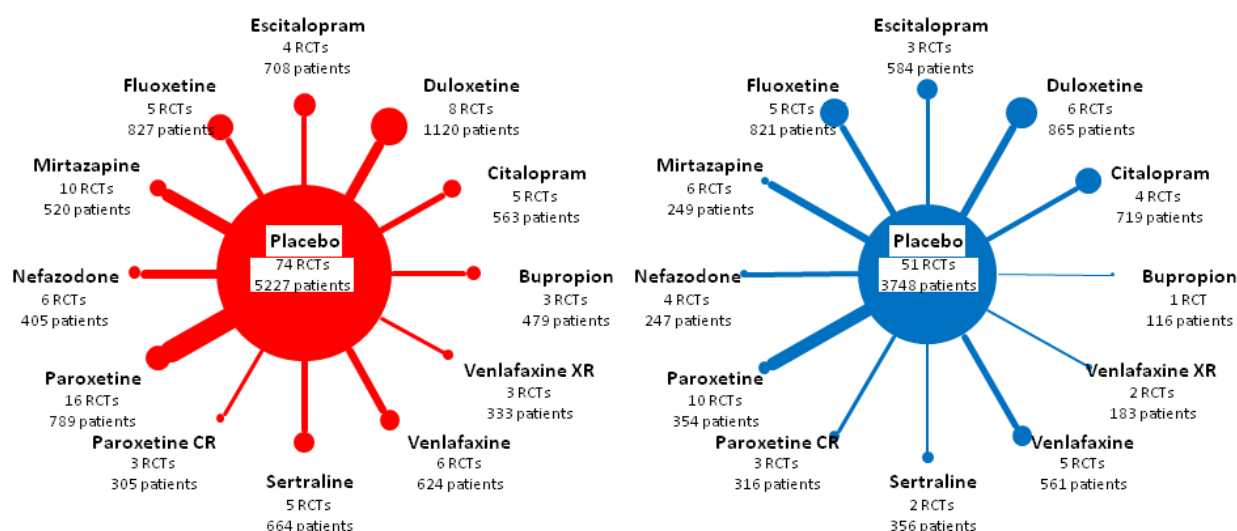
La figure 3.1 montre les 2 réseaux en étoile, avec le placebo en leurs centres, pour les 74 essais enregistrés par la FDA et les 51 essais avec résultats publiés. La proportion d'essais avec des résultats non publiés variait entre les agents antidépresseurs, de 0% pour la fluoxétine et la paroxétine CR à 60% et 67% pour la sertraline et le bupropion, respectivement. Des méta-analyses conventionnelles ont montré une efficacité diminuée pour tous les médicaments en passant des données publiées aux données FDA, la diminution de la quantité d'effet allant de 10% et 11% pour la fluoxétine et la paroxétine CR à 39% et 41% pour la mirtazapine et la nefazodone (Tableau 3.1). L'inspection visuelle des graphiques en entonnoir pour les 12 comparaisons entre chaque médicament et placebo sur les données publiées n'a révélé aucune asymétrie (Annexe 1, Figure 1).

Tableau 3.1: Quantités d'effet des 12 antidépresseurs (relativement au placebo) estimées par méta-analyse conventionnelle

Agent	Données FDA			Données publiées			Données FDA non publiées		
	N	SMD (IC95%)	τ^2	N	SMD (IC95%)	τ^2	N	SMD (IC95%)	τ^2
Bupropion	3	0.17 [0.04; 0.31]	0.00	1	0.27 [0.01; 0.53]	NA	2	0.14 [-0.02; 0.30]	NA
Citalopram	5	0.25 [0.10; 0.38]	0.00	4	0.30 [0.16; 0.44]	0.00	1	0.01 [-0.28; 0.30]	NA
Duloxetine	8	0.30 [0.21; 0.40]	0.00	6	0.40 [0.29; 0.51]	0.00	2	0.15 [-0.05; 0.35]	NA
Escitalopram	4	0.31 [0.18; 0.44]	0.00	3	0.36 [0.23; 0.48]	0.00	1	0.15 [-0.10; 0.39]	NA
Fluoxetine	5	0.26 [0.06; 0.45]	0.02	5	0.29 [0.01; 0.49]	0.02	0	-	-
Mirtazapine	10	0.35 [0.17; 0.54]	0.04	6	0.57 [0.39; 0.75]	0.00	4	0.19 [-0.17; 0.56]	0.09
Nefazodone	6	0.26 [0.12; 0.40]	0.00	4	0.44 [0.26; 0.61]	0.00	2	0.09 [-0.17; 0.34]	0.00
Paroxetine	16	0.42 [0.30; 0.54]	0.00	10	0.59 [0.44; 0.74]	0.00	6	0.20 [-0.00; 0.39]	0.00
Paroxetine CR	3	0.32 [0.15; 0.49]	0.00	3	0.36 [0.20; 0.51]	0.00	0	-	-
Sertraline	5	0.26 [0.12; 0.39]	0.00	2	0.42 [0.24; 0.60]	0.00	3	0.18 [-0.05; 0.40]	0.00
Venlafaxine	6	0.40 [0.24; 0.55]	0.01	5	0.51 [0.36; 0.65]	0.00	1	0.11 [-0.21; 0.44]	NA
Venlafaxine XR	3	0.40 [0.18; 0.62]	0.02	2	0.51 [0.30; 0.71]	0.00	1	0.19 [-0.08; 0.46]	NA

Pour chaque agent antidépresseur, une quantité d'effet moyenne a été estimée par une méta-analyse à effets aléatoires (méthode de DerSimonian et Laird). N: nombre d'essais; SMD (95%IC): différence moyenne standardisée estimée par la méta-analyse (intervalle de confiance à 95%); τ^2 : variance inter-essais (comme mesure de l'hétérogénéité statistique); NA: non estimable.

Figure 3.1: Réseaux d'essais de 12 antidépresseurs issus de 74 essais enregistrés par la FDA et des 51 publications associées

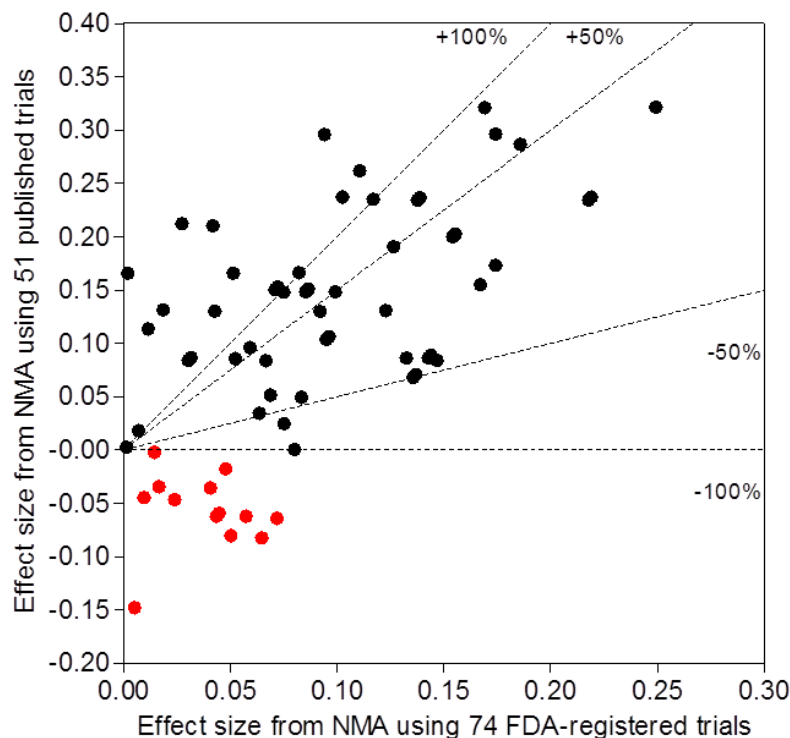


Le nœud central représente le placebo, et chaque nœud externe représente un agent antidépresseur. Le diamètre de chaque nœud est proportionnel au nombre de patients qui ont reçu l'agent antidépresseur; l'épaisseur de chaque ligne est proportionnelle au nombre d'essais qui portaient sur la comparaison.

3.3.1 Impact des biais de publication

La figure 3.2 montre le nuage de points des quantités d'effet pour les 66 comparaisons deux à deux possibles entre antidépresseurs obtenus par la méta-analyse en réseau des 51 essais publiés contre ceux obtenus par la méta-analyse en réseau des 74 essais enregistrés par la FDA. Les estimations différaient en valeur absolue d'au moins 100% dans 30 des 66 comparaisons deux à deux (45%) et d'au moins 50% dans 44 (67%). La différence médiane relative entre les quantités d'effet des 2 méta-analyses en réseau était de 86,1% (25% - 75% percentile 41,4% - 203,8%). Nous avons trouvé 18 quantités d'effet d'intensité modérée (0,2-0,5) à partir des données publiées mais seulement 3 avec les données de la FDA. L'agent B était supérieur à l'agent A à partir des données publiées alors que A était supérieur à B à partir des données FDA dans 13 comparaisons (20%). La significativité statistique était atteinte pour 9 des 66 comparaisons à partir des données publiées et pour seulement 2 comparaisons à partir des données FDA. Les résultats détaillés sont rapportés en Annexe 1, Tableau 1.

Figure 3.2: Quantités d'effet pour les 66 comparaisons deux à deux possibles entre les 12 antidépresseurs, estimées par la méta-analyse en réseau des 74 essais enregistrés par la FDA et la méta-analyse en réseau des 51 essais avec résultats publiés

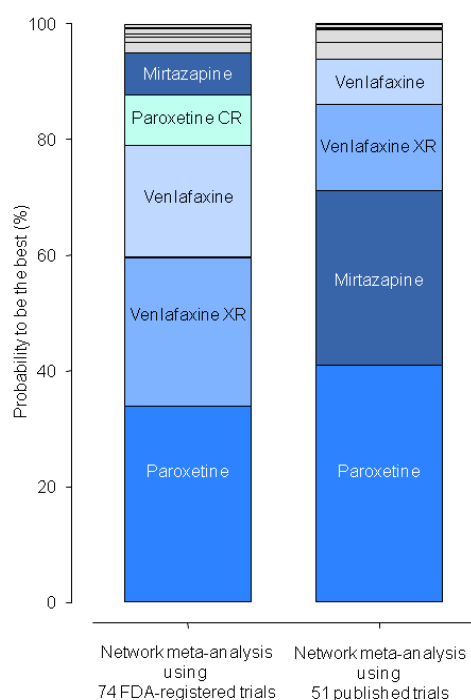


Les données sont des quantités d'effet résumées. Une quantité d'effet positive indique que l'agent A est plus efficace que l'agent B. L'aire au-dessus de la ligne pointillée supérieure (étiquetée +100%) et l'aire en-dessous de la ligne pointillée inférieure (étiquetée -100%) correspondent aux cas où la quantité d'effet estimée via la méta-analyse en réseau des 51 essais avec résultats publiés différait en valeur absolue de celle via la méta-analyse en réseau des 74 essais enregistrés par la FDA d'au moins 100%. L'aire entre les lignes pointillées supérieures (étiquetée +50%) et l'aire entre les lignes pointillées inférieures (étiquetée -50%) correspondent aux cas où la quantité d'effet estimée via la méta-analyse en réseau des 51 essais avec résultats publiés différait en valeur absolue de celle via la méta-analyse en réseau des 74 essais enregistrés par la FDA d'au moins 50%. Les points rouge correspondent aux cas où l'agent B était supérieur à l'agent A à partir des données publiées alors que A était supérieur à B à partir des données FDA.

La figure 3.3 résume les probabilités d'être le meilleur antidépresseur. Ces probabilités varient selon que la méta-analyse en réseau soit basée sur les données publiées ou les données de la FDA: 30,2% ou 7,3% pour la mirtazapine, 41,0% ou 33,9% pour la paroxétine, 0,2% ou 8,7% pour la paroxétine CR, 7,7% ou 19,3% pour la venlafaxine, 14,9% ou 25,7% pour la venlafaxine XR. Ces probabilités allaient de 0% à 3,0% pour les autres agents et selon les données utilisées. De plus, les 3 premiers agents différaient selon les jeux de données utilisés. Dans la méta-analyse en réseau des 51 essais publiés, la paroxétine et la mirtazapine occupaient la première place à égalité alors que la venlafaxine et la venlafaxine XR étaient à

égalité au troisième rang; dans la méta-analyse en réseau des 74 essais enregistrés par la FDA, la paroxétine était la première, et la venlafaxine et la venlafaxine XR étaient à égalité au deuxième rang. La paroxétine est au premier rang dans les deux analyses, et la mirtazapine a été propulsée en avant dans le classement à l'issue de l'analyse des données publiées. Des graphiques et des métriques complémentaires, notamment les courbes des probabilités de rang (en anglais, rankogram) et les surfaces sous la courbe des probabilités de rang cumulées pour chaque traitement sont rapportés en Annexe 1, Figure 2 à 5.

Figure 3.3: Probabilité que chaque antidépresseur soit le meilleur d'après la méta-analyse en réseau des 74 essais enregistrés par la FDA et la méta-analyse en réseau des 51 essais avec résultats publiés



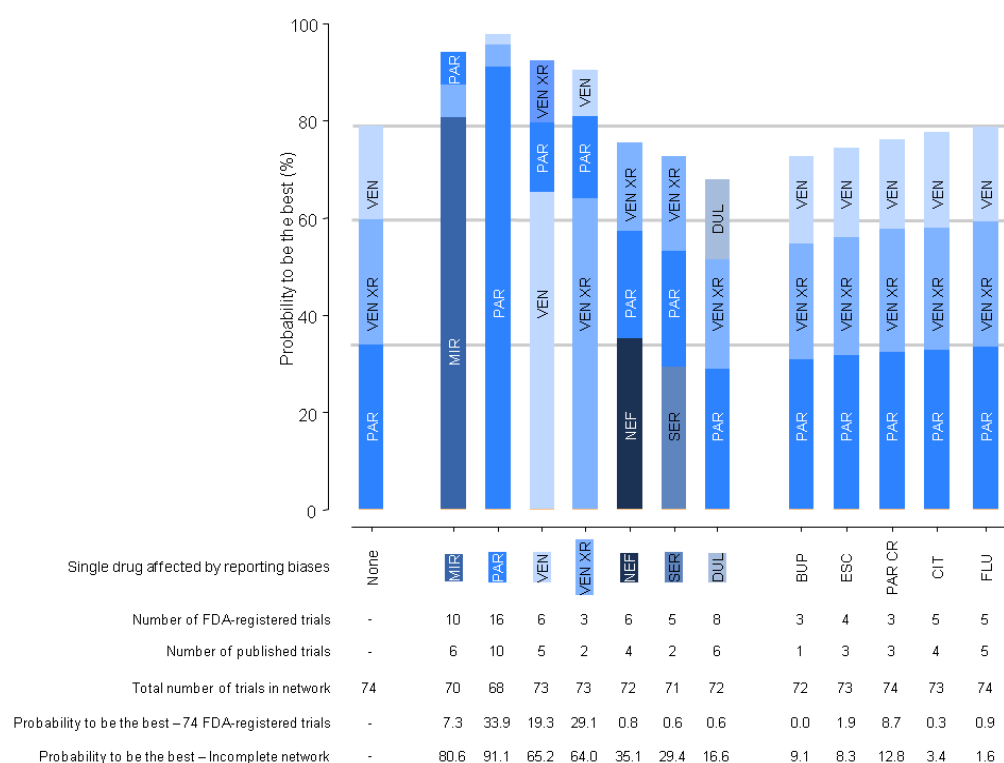
Par exemple, la probabilité que la mirtazapine soit le meilleur agent était de 7.3% et 30.2% d'après la méta-analyse en réseau des 74 essais enregistrés par la FDA et la méta-analyse en réseau des 51 essais avec résultats publiés, respectivement. Les agents pour lesquels la probabilité d'être le meilleur était < 5% d'après les 2 analyses sur données publiées et données de la FDA ne sont pas étiquetés (aires grisées).

3.3.2 Impact du biais de publication affectant un seul médicament

La figure 3.4 montre les résultats de la méta-analyse en réseau en supposant que le biais de publication n'affectait qu'un seul médicament (en utilisant les essais avec résultats

publiés pour ce médicament et les essais enregistrés par la FDA pour les 11 autres agents). Par exemple, nous avons utilisé les 6 essais avec résultats publiés pour la mirtazapine (alors que 10 essais étaient enregistrés par la FDA) et les 64 essais enregistrés par la FDA pour les 11 autres agents, ce qui donne un réseau incomplet de 70 essais; la probabilité que la mirtazapine soit le meilleur antidépresseur valait 80.6% en analysant le réseau incomplet and 7.3% en analysant le réseau complet de 74 essais enregistrés par la FDA.

Lorsqu'un seul agent était hypothétiquement affecté par les biais de publication, l'agent concerné a été dans la plupart des cas fortement favorisé. L'augmentation de la probabilité d'être le meilleur à partir de la méta-analyse en réseau du réseau incomplet plutôt que des 74 essais enregistrés par la FDA variait de 0,7% à 73,3%. L'agent affecté par le biais de publication était au premier rang dans 5 des 12 analyses, mais seconde ($n = 2$), quatrième ($n = 1$) ou huitième ($n = 2$) dans la méta-analyse en réseau des 74 essais enregistrés par la FDA. En outre, le classement des autres médicaments pouvait être modifié. Dans 6 des 12 analyses, les 3 premiers agents différaient de ceux obtenus via l'analyse des données de la FDA, tandis que seules de petites modifications au classement des médicaments sont survenues dans 5 autres des 12 analyses.

Figure 3.4: Probabilités d'être le meilleur antidépresseur lorsque le biais de publication n'affecte qu'un seul agent

La première barre à gauche correspond à la méta-analyse en réseau des données sans biais de publication (i.e., les 74 essais enregistrés par la FDA). Les autres barres correspondent aux 12 méta-analyses en réseau des données hypothétiques où le biais de publication affecte un agent spécifique. Par exemple, nous avons utilisé les 6 essais avec résultats publiés pour la mirtazapine (alors que 10 essais sont enregistrés par la FDA) et les 64 essais enregistrés par la FDA pour les 11 autres agents, ce qui donne un réseau incomplet de 70 essais; la probabilité que la mirtazapine soit le meilleur antidépresseur valait 80.6% en analysant le réseau incomplet et 7.3% en analysant le réseau complet de 74 essais. Dans un souci de clarté, nous avons présenté uniquement les 3 agents avec les probabilités les plus élevées pour chaque analyse. Bup: Bupropion; Cit: Citalopram; Dul : Duloxetine; Esc: Escitalopram; Flu: Fluoxetine; Mir: Mirtazapine; Nef: Nefazodone; Par: Paroxetine; Par CR: Paroxetine CR; Ser: Sertraline; Ven: Venlafaxine; VenXR: Venlafaxine XR.

3.3.3 Impact spécifique du biais de publication et de la présentation sélective des analyses

Pour le biais de publication spécifiquement, les quantités d'effet obtenues à partir de la méta-analyse en réseau des 51 essais publiés (mais avec les données de la FDA) et celles obtenues à partir de la méta-analyse en réseau des 74 essais enregistrés par la FDA différaient d'au moins 100% dans 19 des 66 comparaisons (29%) et d'au moins 50% dans 36 (55%). La différence médiane relative entre les quantités d'effet deux à deux à partir de ces 2 analyses était de 60,6% (25% - 75% percentile 28,5% - 103,3%).

De même, pour la présentation sélective des analyses, les quantités d'effet obtenues à partir de la méta-analyse en réseau des 51 essais publiés (avec les quantités d'effet publiées) et celles obtenues à partir de la méta-analyse en réseau des 51 essais publiés mais avec les quantités d'effet de la FDA différaient d'au moins 100% dans 21 des 66 comparaisons (32%) et d'au moins 50% dans 35 (53%). La différence médiane relative entre les quantités d'effet deux à deux à partir de ces 2 analyses était de 56,2% (25% - 75% percentile 16,3% - 135,7%).

3.4 Discussion

Dans cette étude, nous avons évalué l'impact des biais de publication sur les résultats de la méta-analyse en réseau, en prenant l'exemple des essais randomisés contre placebo de 12 antidépresseurs enregistrés par la FDA et leurs publications correspondantes. Premièrement, nous avons constaté d'importantes différences dans les estimations de l'efficacité relative des antidépresseurs concurrents obtenues à partir des méta-analyses des données de la FDA et des données publiées. Pour environ la moitié des comparaisons deux à deux entre médicaments, les quantités d'effet issues de la méta-analyse en réseau des données publiées différaient, en valeur absolue, d'au moins 100% de celles issues de la méta-analyse en réseau des données de la FDA. Le classement des agents selon leur efficacité a également été influencé, avec des différences dans la probabilité d'être le meilleur agent. Deuxièmement, les biais de publication affectant un seul médicament peuvent influencer le classement de toutes les molécules. Troisièmement, le biais de publication spécifiquement mais également la présentation sélective des analyses statistiques contribuent à ces résultats.

Notre étude, basée sur les essais d'antidépresseurs enregistrés par la FDA et leurs publications correspondantes, n'avait pas pour but de comparer les antidépresseurs les uns aux autres, mais plutôt d'évaluer l'impact des biais de publication sur les méta-analyses en réseau. Nous avons utilisé ce jeu de données déjà décrit et publié précédemment par Turner et al.

(Turner, Matthews et al. 2008), car à notre connaissance, c'est le seul jeu de données disponible qui offre la possibilité d'évaluer l'impact des biais de publication sur un réseau d'essais. D'autres études ont comparé essais enregistrés par la FDA et résultats publiés, mais ces études ne concernaient pas tous les médicaments concurrents pour une condition spécifique et ne permettaient donc pas d'effectuer une méta-analyse en réseau (Lee, Bacchetti et al. 2008; Rising, Bacchetti et al. 2008).

Notre étude apporte 3 nouvelles informations. Premièrement, notre analyse a porté sur les méta-analyses en réseau. Une littérature abondante a montré l'existence et l'impact des biais de publication dans les méta-analyses conventionnelles, y compris l'étude de Turner et al. (Turner, Matthews et al. 2008). Cependant, cette question est encore mal explorée dans la comparaison indirecte ajustée ou dans le cadre des méta-analyses en réseau. En particulier, les méta-analyses en réseau actuelles n'abordent pas ou peu la possibilité des biais de publication. Dans cette étude de cas, nous avons montré que la synthèse des données publiées conduirait à des estimations erronées de l'efficacité des interventions concurrentes en présence de biais de publication. Si l'on dispose d'une preuve ou d'un signal indiquant la présence de biais de publication pour l'une des méta-analyses conventionnelles constituant le réseau, les résultats de la méta-analyse en réseau doivent être interprétés avec beaucoup de prudence. Mais la reconnaissance de ce problème est d'autant plus important qu'il n'existe pas actuellement de méthode reconnue pour identifier et/ou corriger le biais de publication dans le cadre des méta-analyses en réseau. On pourrait certes construire des graphiques en entonnoir (funnel plot) et appliquer des tests d'asymétrie pour chaque comparaison deux à deux dans le réseau. Toutefois, le nombre d'essais pour chaque comparaison deux à deux est souvent très limité, ce qui ne permettrait pas à cette approche de documenter ou d'exclure l'existence d'un biais de manière appropriée (Ioannidis and Trikalinos 2007; Moreno, Sutton et al. 2009). Dans notre exemple, chacune des 12 comparaisons était représentée par moins de 10 essais, donc

l'application de tests d'asymétrie serait inappropriée (Sterne, Sutton et al. 2011). En outre, même en connaissant l'existence d'essais enregistrés par la FDA et non publiés, l'évaluation visuelle des graphiques en entonnoir des données publiées n'a révélé aucune asymétrie (Annexe 1, Figure 1). Dans certains contextes, on pourrait supposer que les biais de publication affectent les différents médicaments de manière similaire et supposer l'interchangeabilité des processus de sélection d'essais entre médicaments; des méthodes qui tirent leur force de tous les essais du réseau pourraient être appliquées, comme cela a été réalisé récemment pour l'étude de cas que nous avons considérée (Moreno, Sutton et al. 2009; Moreno, Sutton et al. 2011). De plus, dans certaines situations, il n'est probablement pas nécessaire que le biais de publication soit intense pour influencer les résultats. Par exemple, les biais de publication affectant les essais venlafaxine sont liés à seulement un essai avec des résultats non publiés sur un total de 6 essais enregistrés par la FDA; lorsque, de manière hypothétique, le biais de publication affectait la venlafaxine seule, la venlafaxine apparaissait au premier rang.

Deuxièmement, nous avons également montré que le biais de publication fonctionne différemment dans la méta-analyse en réseau et dans la méta-analyse conventionnelle. La principale différence est que, dans une méta-analyse conventionnelle, le biais de publication ne concerne que les résultats de l'intervention d'intérêt. En revanche, dans la méta-analyse en réseau, les biais de publication affectant un traitement parmi plusieurs peuvent affecter le classement de tous les traitements (« une brebis galeuse gâte tout le troupeau »). Les résultats d'une méta-analyse en réseau ne peuvent être valables que si les résultats de recherches menées indépendamment sont rendus disponibles de manière équiprobable.

Troisièmement, le biais de publication spécifiquement mais également la présentation sélective des analyses statistiques ont influencé les estimations de l'efficacité relative des traitements. L'impact des biais sur la méta-analyse en réseau n'est pas la somme algébrique de

l'impact des deux types de biais. De surcroît, l'impact considérable du biais de présentation sélective des analyses est d'une importance pratique majeure. En effet, contrairement au biais de publication, la présentation sélective des analyses est moins susceptible d'être améliorée par le seul enregistrement des essais (e.g., clinicaltrials.gov), mais une analyse statistique indépendante supplémentaire avec accès à l'ensemble des données brutes, au protocole d'essai et au plan d'analyse pré-spécifiée pourrait être nécessaire pour assurer l'intégrité (Zarin and Tse 2008; DeAngelis and Fontanarosa 2010). Dans tous les cas, les méta-analyses en réseau comme les méta-analyses conventionnelles devraient inclure les données issues des rapports d'évaluation des médicaments approuvés par la FDA, qui sont désormais accessibles au public à partir du website du Centre for Drug Evaluation and Research (Turner 2005; O'Connor 2009).

Notre étude présente plusieurs limites. Premièrement, cette analyse empirique s'appuie sur un réseau particulier dédié à une condition clinique particulière (la dépression majeure), à une classe de médicaments (antidépresseurs) et un type d'essais (essais randomisés contre placebo). Cette spécificité pourrait limiter la généralisation de nos résultats. Dans cette étude de cas, les biais de publication ont conduit à une surestimation de l'effet pour tous les médicaments. Cependant, une nouvelle analyse de méta-analyses avec l'ajout de données non publiées de la FDA pour 6 autres classes de médicaments a montré que l'effet de cet ajout de données non publiées de la FDA variait selon le médicament et le critère de jugement, avec la possibilité d'une diminution, mais aussi d'une augmentation de l'efficacité estimée (Hart, Lundh et al. 2012). Le réseau d'essais de Turner et al. était limité aux essais randomisés contre placebo, ce qui entraîne une forme d'étoile rayonnante. Toutefois, elle correspond à la géométrie des réseaux existants, puisque les réseaux simples avec 3 interventions différentes sans essai face-face sont fréquents (60% des comparaisons indirectes identifiées dans une revue récente (Donegan, Williamson et al. 2010)) et, dans le cas de trois interventions ou plus,

des exemples de réseaux avec des motifs en étoile ou en échelle (ne contenant pas de boucle) ont été fréquemment rapportés (Salanti, Kavvoura et al. 2008; O'Regan, Ghement et al. 2009). De plus, le réseau de Turner et al ne comprenaient pas les essais face-face entre ces antidépresseurs qui existent pourtant (Gartlehner, Gaynes et al. 2008; Cipriani, Furukawa et al. 2009). Nous n'avons pas accès aux essais face-face dont les résultats n'ont pas été publiés. En effet, l'enregistrement réglementaire est rare pour les essais face-face et si des données fournies par l'industrie étaient disponibles, elles pourraient encore souffrir d'un biais de présentation sélective des critères de jugement ou des analyses. Par conséquent, nous n'avons pas pu effectuer une méta-analyse en réseau sur un réseau complet et non biaisé (i.e., incluant tous les essais randomisés contre placebo et face-face). Cependant, l'ajout à notre analyse d'essais face-face avec résultats publiés ou non publiés pourrait modifier l'impact estimé des biais de publication. L'extrapolation de nos résultats à un réseau d'essais avec à la fois de l'information directe et indirecte n'est pas simple. La direction du biais dans l'estimation de l'effet du traitement en raison des biais de publication est incertaine pour les essais face-face: le traitement produit par le promoteur de l'essai pourrait être favorisé (Lexchin, Bero et al. 2003; Lathyris, Patsopoulos et al. 2010) ou le traitement plus récent pourrait être favorisé (Chalmers and Matthews 2006; Bero, Oostvogel et al. 2007). De surcroît, distinguer les sources de biais affectant l'information directe et indirecte serait difficile.

Deuxièmement, le choix des essais enregistrés par la FDA comme norme de référence pourrait être discuté, mais il semble raisonnable. Les quantités d'effet deux à deux dérivées de l'analyse des données de la FDA ne devraient pas être considérées comme des estimations sans biais de l'efficacité *per se* des antidépresseurs, mais peuvent être considérées comme des estimations non biaisées de l'effet des traitements via la méta-analyse en réseau des essais randomisés contre placebo. En effet, au cours du processus d'examen des demandes d'enregistrement de nouveaux médicaments, la FDA ré-analyse les données des essais, en

utilisant les données brutes fournies par le promoteur, dans le respect des méthodes statistiques préétablies dans le protocole (Turner 2004). Le jeu de données de la FDA que nous avons utilisé a déjà été décrit comme « unbiased (but not the complete) body of evidence » (Moreno, Sutton et al. 2009).

Troisièmement, la vérification des hypothèses requises pour la méta-analyse en réseau est, comme toujours, difficile. Cependant, il n'y a aucune raison de penser qu'ici ces conditions ne sont pas remplies. L'hypothèse d'homogénéité était vérifiée dans notre analyse. La similarité des essais est très vraisemblable, parce que tous les essais étaient des essais randomisés contre placebo, en double aveugle, sur des médicaments pour le traitement à court terme de la dépression, avec des critères de sélection proches. De plus, d'autres méta-analyses en réseau ont été réalisées dans ce domaine et n'ont pas soulevé de problèmes au sujet de ces hypothèses (Gartlehner, Gaynes et al. 2008; Cipriani, Furukawa et al. 2009). En outre, si l'une de ces hypothèses n'avait pas été respectée, notre analyse n'aurait probablement pas été influencée, car cela aurait sans doute concerné les deux méta-analyses en réseau des données publiées et des données de la FDA que nous avons comparées. Une hypothèse supplémentaire requise pour les méta-analyses en réseau est l'interchangeabilité, ce qui implique que si tous les essais avait inclus tous les traitements évalués dans le réseau, chaque essai aurait estimé les mêmes quantités d'effet. L'hypothèse de cohérence est strictement dérivée de l'hypothèse d'interchangeabilité. Les réseaux en étoile ne permettent pas de quantifier l'incohérence entre information directe et indirecte. La disponibilité inégale des essais pour différentes comparaisons, en raison de biais de publication, peut entraîner l'incohérence (par exemple, des biais de publication différentiels d'un agent à l'autre peuvent conduire à la violation de cette hypothèse) (Salanti, Higgins et al. 2008). Lorsque le biais de publication affectait de manière hypothétique un seul médicament, nous avons essentiellement examiné les conséquences de la violation du principe de l'interchangeabilité et nous avons constaté que le

classement de tous les médicaments pouvait être modifié. De tels biais de publication différentiels pourraient se produire entre interventions concurrentes mais également au sein des interventions concurrentes. Par exemple, les biais de publication pourraient différer entre les essais réalisés avant et après le FDA Amendments Act de 2007, qui a étendu les exigences légales pour les rapports d'essai (Miller 2010).

3.5 Conclusion

La méta-analyse en réseau est un outil statistique prometteur, en particulier pour la Comparative Effectiveness Research, mais les auteurs doivent être conscients de l'impact potentiel des biais de publication sur les résultats d'une telle synthèse. La validité d'une méta-analyse en réseau est conditionnée par la disponibilité égale des résultats de recherches menées indépendamment. Les auteurs doivent interpréter les résultats avec beaucoup de prudence lorsqu'un biais de publication est détecté dans une comparaison deux à deux donnée et, doivent être surtout conscients que les biais de publication ne sont probablement pas détectés ni exclus de manière appropriée dans le cadre des méta-analyses en réseau.

Chapitre 4

Détection du biais de publication dans la méta-analyse en réseau

4.1 Contexte

4.2 Test d'excès de résultats significatifs

4.3 Simulation

4.4 Résultats

4.5 Exemple d'application dans un réseau d'essais d'antidépresseurs

4.6 Discussion

4.7 Conclusion

4.1 Contexte

Le biais de publication est la tendance qu'ont les chercheurs, les laboratoires pharmaceutiques et les revues à publier les résultats d'essais selon la direction, l'intensité et la significativité statistique des résultats (McGauran, Wieseler et al. 2010; Song, Parekh et al. 2010). Ce biais est encore un sujet de préoccupation majeur pour les méta-analystes, car si les résultats « positifs » sont plus susceptibles d'être disponibles, une méta-analyse basée sur ces résultats sera biaisée (Hopewell, Loudon et al. 2009). En outre, la présentation sélective de certaines analyses statistiques et de certains critères de jugement est de plus en plus considérée comme plus fréquente et donc comme une menace potentiellement plus grande que le biais de publication spécifiquement (Dwan, Altman et al. 2008).

De nombreux tests statistiques ont été proposés pour détecter la présence de biais dans une méta-analyse conventionnelle (Begg and Mazumdar 1994; Macaskill, Walter et al. 2001; Schwarzer, Antes et al. 2002; Harbord, Egger et al. 2006; Rucker, Schwarzer et al. 2008; Song, Parekh et al. 2010; Rucker, Carpenter et al. 2011). La plupart d'entre eux évaluent

formellement le degré d'asymétrie dans le graphique en entonnoir comme un signe de l'effet des études de petite taille, c'est-à-dire la tendance qu'ont les petits essais à montrer des estimations de l'effet du traitement plus grandes, quelles que soient les raisons de cet effet (Ioannidis and Trikalinos 2007; Sterne, Sutton et al. 2011). Au lieu de tester si les petits essais ont donné des estimations différentes de celles des grands essais, Ioannidis et Trikalinos ont proposé une approche différente en recherchant un excès d'essais avec des résultats statistiquement significatifs dans une méta-analyse (Ioannidis and Trikalinos 2007). Contrairement à d'autres tests, les performances de ce test n'ont pas été évaluées par des études de simulation sous différents scénarios, et la Collaboration Cochrane n'a pas encore recommandé son utilisation pour les méta-analyses conventionnelles (Sterne, Egger et al. 2011). Le test est peu puissant lorsqu'il y a peu d'essais avec résultats statistiquement significatifs dans une méta-analyse. Cependant, il présente cet avantage considérable de pouvoir être utilisé pour identifier des biais affectant de plus grands ensembles de données issues de plusieurs méta-analyses. Le test a ainsi été utilisé pour l'évaluation du biais sur des domaines entiers de recherche, en génétique, neurologie ou sur les biomarqueurs, à partir des données de plusieurs dizaines à plusieurs centaines de méta-analyses (Kavvoura, McQueen et al. 2008; Chatzinasiou, Lill et al. 2011; Ioannidis 2011; Tsilidis, Papatheodorou et al. 2012; Tzoulaki, Siontis et al. (in press)).

La situation classique où l'on réunit plusieurs méta-analyses d'essais randomisés portant sur la même maladie et le même critère de jugement est la méta-analyse en réseau (Salanti, Higgins et al. 2008; Ioannidis 2009; Salanti, Ades et al. 2010). Les méta-analyses conventionnelles qui sont combinées dans une méta-analyse en réseau évaluent des comparaisons entre différents traitements, mais les essais sont considérés comme suffisamment similaires pour être synthétiser dans le même réseau. Le biais de publication est également un problème majeur, sans doute négligé, pour la méta-analyse en réseau (Li, Puhan

et al. 2011; Mills, Ioannidis et al. 2012). Bien que cette méthode soit de plus en plus utilisée, le biais de publication a reçu peu d'attention dans le cadre de la méta-analyse en réseau. Alors que les biais de publication peuvent avoir un impact important sur les estimations de l'efficacité des traitements, de nombreuses méta-analyses en réseau publiées n'ont pas abordé la possibilité d'un biais de publication et de biais liés aux petites études (Bafeta, Trinquart et al. ; Trinquart, Abbé et al. 2012). En outre, il n'existe aucune méthode pour la détection des biais de publication dans la méta-analyse en réseau. Un test qui fournirait un signal statistique, à condition qu'il soit utilisé judicieusement, complèterait et renforcerait la méta-analyse en réseau comme un effort systématique visant à résumer les données provenant d'un réseau d'essais. Nous proposons d'étendre le test proposé par Ioannidis et Trikalinos à la méta-analyse en réseau. Ce test a été introduit à l'origine pour évaluer la présence de biais sur des domaines de recherche entiers et le test peut être appliqué à un réseau en tirant partie de tous les essais (Pan, Trikalinos et al. 2005; Ntzani, Rizos et al. 2007; Kavvoura, McQueen et al. 2008; Ioannidis 2011; Tsilidis, Papatheodorou et al. 2012).

Dans la section 4.2, nous décrivons le test dans une méta-analyse conventionnelle et le test généralisé dans une méta-analyse en réseau. Dans la section 4.3, nous décrivons le protocole d'études de simulation pour évaluer la performance statistique des tests selon divers scénarios. Les résultats sont présentés dans la section 4.4. La section 4.5 montre deux exemples d'application sur des réseaux d'essais d'antidépresseurs et d'antipsychotiques, sur lesquels l'existence de biais de publication, de forte et faible intensité, respectivement, a déjà été démontrée par la comparaison de résultats publiés à de la FDA. Nous concluons par une discussion dans la section 4.6.

4.2 Test d'excès de résultats significatifs

4.2.1 Détection du biais dans une méta-analyse conventionnelle

Nous considérons une méta-analyse de N_0 essais. Nous testons si le nombre observé O d'essais avec des résultats statistiquement significatifs à un niveau α spécifié est différent du nombre attendu E d'essais avec des résultats statistiquement significatifs. Nous fixons $\alpha = 0,05$.

Pour estimer le nombre attendu E , nous supposons qu'il existe un effet du traitement vrai θ . La probabilité attendue qu'un essai détecte une différence statistiquement significative correspond à sa puissance $1 - \beta_i$ pour détecter θ au niveau α spécifié. Le nombre attendu E est égal à la somme des probabilités attendues sur tous les essais sélectionnés dans la méta-analyse.

Pour toute méta-analyse, la quantité d'effet vraie θ n'est pas connue. Par conséquent, nous devons faire une hypothèse sur un effet traitement plausible. Cette quantité d'effet plausible peut être: l'effet combiné obtenu via une méta-analyse à effet fixe, l'effet combiné obtenu via une méta-analyse à effets aléatoires ou l'estimation de l'effet du traitement fournie par la plus grande étude dans la méta-analyse, respectivement. Plutôt que de considérer une estimation ponctuelle, on pourrait tenir compte de la distribution a priori de l'effet traitement plausible. Cependant, les résultats obtenus dans ce cas sont similaires à ceux obtenus en considérant l'estimation ponctuelle, ce que nous utilisons par la suite (Ioannidis and Trikalinos 2007).

On teste si le nombre observé O est supérieur au nombre E à un niveau α' spécifié par un test de probabilité binomiale. Nous fixons $\alpha'=0,10$. Par conséquent, nous rejetons l'hypothèse nulle en faveur d'un excès de résultats significatifs si $\sum_{i=0}^N b(i|N_0, E/N_0) < \alpha'$, où $b(x|N_0, p) = \binom{N_0}{x} p^x (1-p)^{N_0-x}$ est la densité binomiale. Comme le test est basé sur les

nombre attendus et observés de résultats significatifs, il est applicable à tout type de critère de jugement, entre autres aux données binaires, continues ou de survie.

Pour des données binaires, admettons que nous observions x_{Ei} et x_{Ci} événements parmi n_{Ei} et n_{Ci} patients dans les groupes expérimental et témoin de l'essai i , respectivement. À la manière de Ioannidis et Trikalinos, nous estimons la proportion d'événements vraie dans le groupe témoin par $\hat{\pi}_{Ci}=x_{Ci}/n_{Ci}$ et la proportion d'événements vraie dans le groupe expérimental par $\hat{\pi}_{Ei}=[1+(n_{Ci}-x_{Ci})/(\theta \cdot x_{Ci})]^{-1}$, avec θ le vrai rapport de cotes entre les groupes expérimental et témoin. Ensuite, $1 - \beta_i$ est estimé comme étant la puissance du test bilatéral exact de Fisher pour détecter la différence entre $\hat{\pi}_{Ei}$ et $\hat{\pi}_{Ci}$ parmi n_{Ei} et n_{Ci} patients au niveau spécifié α . Enfin, O est le nombre d'essais avec valeur p significative au niveau spécifié α avec un test bilatéral exact de Fisher sur x_{Ei} , x_{Ci} , n_{Ei} , n_{Ci} .

Pour des données continues, admettons que nous observions les moyennes et les écarts types m_{Ei} , s_{Ei} et m_{Ci} , s_{Ci} de n_{Ei} et n_{Ci} patients dans les groupes expérimental et témoin de l'essai i , respectivement. La moyenne vraie dans le groupe témoin est estimée par $\hat{\mu}_{Ci} = m_{Ci}$ et la moyenne vraie dans le groupe expérimental est estimée par $\hat{\mu}_{Ei} = \hat{\mu}_{Ci} + s_i \cdot g$ avec g la différence moyenne standardisée vraie entre les groupes expérimental et témoin et $s_i = \sqrt{((n_{Ei} - 1)s_{Ei}^2 + (n_{Ci} - 1)s_{Ci}^2)/(n_{Ei} + n_{Ci} - 2)}$ l'écart type intra-groupes. Ensuite, $1 - \beta_i$ est estimée comme étant la puissance du test bilatéral de Student pour détecter la différence entre $\hat{\mu}_{Ei}$ et $\hat{\mu}_{Ci}$ sur n_{Ei} et n_{Ci} au niveau spécifié α . Enfin, O est le nombre d'essais avec valeur p significative au niveau spécifié α avec un test bilatéral de Student basé sur m_{Ei} , s_{Ei} , n_{Ei} et m_{Ci} , s_{Ci} , n_{Ci} .

4.2.2 Détection du biais dans une méta-analyse en réseau

Nous considérons qu'un réseau d'essais peut être décrit comme J méta-analyses de N_j essais chacune. Nous estimons le nombre attendu E_j d'essais avec résultats statistiquement

significatifs pour chaque méta-analyse sur le réseau en supposant qu'il y ait une quantité d'effet vraie θ_j pour chaque méta-analyse. Le nombre attendu E d'essais avec résultats statistiquement significatifs sur l'ensemble du réseau est estimé par $E = \sum_{j=1}^J E_j = \sum_{j=1}^J \sum_{i=1}^{N_j} 1 - \beta_i$. La quantité d'effet vraie pour chaque méta-analyse est estimée sur la base d'une quantité d'effet plausible, comme décrit ci-dessus. Le nombre observé $O = \sum_{j=1}^J O_j$ est le nombre total d'essais avec des valeurs p significatives sur l'ensemble du réseau. On teste si le nombre observé O est supérieur au nombre attendu E avec un test de probabilité binomiale. Notons $N = \sum_{j=1}^J N_j$. Nous rejetons l'hypothèse nulle en faveur d'un excès de résultats significatifs si $\sum_{i=0}^N b(i|N, E/N) < \alpha'$.

La détection d'un excès de résultats statistiquement significatifs ne voudrait pas dire que toutes les méta-analyses conventionnelles incluses dans le réseau ont été affectées de manière égale par le même biais. Même si le biais est interchangeable sur le réseau, nous pourrions observer que certaines méta-analyses sont plus affectées que d'autres, et que certaines méta-analyses ne sont pas affectées du tout. Par exemple, pour trois méta-analyses et 10 essais par méta-analyse, si 10% des résultats sont supprimés à cause du biais de publication, il est possible que les résultats de trois essais issus de la même méta-analyse ou de trois essais issus chacun d'une des méta-analyses restent non publiés.

4.3. Simulation

Nous avons évalué les taux d'erreur de type I et de type II des tests à l'aide d'études de simulation de type Monte Carlo. Ces simulations étaient basées sur des données binaires. Premièrement, nous avons évalué le test pour une méta-analyse conventionnelle. Deuxièmement, nous avons évalué le test pour une méta-analyse en réseau.

4.3.1 Méthodes pour générer les données d'un réseau d'essais

4.3.1.1 Simulation des essais dans une méta-analyse

Pour une méta-analyse donnée (i.e., un couple donné de traitements expérimental et témoin), nous avons généré les nombres de patients présentant et ne présentant pas l'événement dans les groupes expérimental et témoin pour chaque essai, dans le cadre d'un modèle à effets aléatoires.

Nous avons fixé l'effet moyen vrai du traitement θ et sa variance inter-essais τ^2 . Admettons que $\theta < 1$ signifie que le traitement expérimental est bénéfique par rapport au traitement témoin.

Nous avons généré les probabilités d'événement vraies dans les groupes expérimental et témoin de l'essai i de la manière suivante: $\text{logit}(p_{Ei}) = \text{logit}(p_i) + \log(\theta) + \tau \cdot \eta_i/2$ et $\text{logit}(p_{Ci}) = \text{logit}(p_i) - \tau \cdot \eta_i/2$, où $\text{logit}(x) = \log[x / (1 - x)]$, avec p_i une réalisation d'une distribution uniforme sur l'intervalle $[0,3; 0,7]$ et η_i une réalisation d'une distribution normale centrée réduite. La distribution du risque d'événement « vrai » p_i pourrait être modifiée pour refléter des risques plus ou moins élevés.

Dans un échantillon de 77237 essais inclus dans 14886 méta-analyses Cochrane avec des données binaires, la médiane de la taille des essais était de 102 et le 75^{ème} percentile était de 243 sujets (Davey, Turner et al. 2011). Par conséquent, nous avons tiré au sort le logarithme du nombre de patients de l'essai i à partir d'une distribution normale avec une moyenne de 4,62 et un écart type de 1,29. Les nombres de patients inférieurs à 20 ont été ramenés à 20. De plus, nous avons arrondi le nombre de patients de l'essai i à l'entier le plus proche, soit $n_i = \lfloor n_i \rfloor + \lfloor n_i \rfloor \bmod 2$, avec $\lfloor x \rfloor$ la partie entière par défaut de x et \bmod l'opération modulo. Enfin, nous avons supposé que le nombre de patients était égal dans les groupes expérimental et de témoin, $n_{Ci} = n_{Ei} = n_i / 2$. Les nombres d'événements x_{Ei} et x_{Ci} dans les groupes expérimental et témoin pour l'essai i ont été tirés au sort à partir d'une distribution

binomiale de paramètres (n_{Ei}, p_{Ei}) et (n_{Ci}, p_{Ci}) , respectivement. Si aucun événement n'était survenu dans les deux groupes, l'essai était exclu et remplacé par un nouvel essai généré. Si aucun événement ne s'était produit dans un seul groupe, un facteur de correction égal à 0,5 a été ajouté à toutes les cellules du tableau de contingence 2×2 pour l'essai correspondant. Ces options pour tenir compte des essais sans événement ou avec une cellule du tableau de contingence 2×2 sont utilisées couramment (Higgins, Deeks et al. 2011).

Avec cette procédure de génération des données, un modèle de méta-analyse à effets aléatoires serait correctement spécifié par

$$\log(OR_i) = \log(\theta_i) + s_i \cdot \varepsilon_i, \varepsilon_i \sim N(0,1) \text{ and } \log(\theta_i) \sim N(\log(\theta), \tau^2) \quad (1)$$

puisque le logarithme du rapport des cotes pour l'essai i peut être estimé par $\log(OR_i) = \log[(x_{Ei}/n_{Ei})/(x_{Ci}/n_{Ci})]$ et sa variance par $s_i^2 = 1/x_{Ei} + 1/(n_{Ei} - x_{Ei}) + 1/x_{Ci} + 1/(n_{Ci} - x_{Ci})$.

4.3.1.2 Simulation du biais de publication

Nous avons utilisé deux modèles de sélection pour simuler les biais de publication affectant les essais d'une méta-analyse. La première relie la probabilité de sélection d'un essai à la fois à la taille de l'essai et à l'intensité de l'effet traitement estimé (Copas and Shi 2000; Copas and Shi 2001; Rucker, Schwarzer et al. 2008; Dias, Welton et al. 2010). La seconde fait dépendre la sélection de l'essai à la valeur p associée à l'effet traitement estimé (Begg and Mazumdar 1994; Macaskill, Walter et al. 2001; Harbord, Egger et al. 2006).

Modèle de sélection en fonction de la taille de l'essai et de l'intensité de l'effet du traitement estimé

Nous avons considéré que le modèle (1) décrivait tous les essais existants, mais que seuls certains d'entre eux étaient sélectionnés dans la méta-analyse avec une propension

$$z_i = a + b/s_i + \delta_i \text{ with } \delta_i \sim N(0,1) \quad (2)$$

Seuls les essais avec une valeur z_i positive sont sélectionnés dans la méta-analyse. Les modèles (1) et (2) sont combinés de sorte que

$$\log(OR_i) = \log(\theta_i) + \sigma_i \cdot \varepsilon_i, \quad \varepsilon_i \sim N(0,1) \text{ with } \log(\theta_i) \sim N(\log(\theta), \tau^2), \quad z_i = a + b/s_i + \delta_i, \quad \delta_i \sim N(0,1)$$

et en supposant que les résidus $(\varepsilon_i, \delta_i)$ suivent une distribution normale bivariée avec les deux moyennes égales à 0, les deux variances égales à 1 et avec une corrélation ρ . L'écart type σ_i de $\log(OR_i)$ diffère de s_i qui est une estimation de la variance conditionnelle de $\log(OR_i)$ sachant que l'essai i a été sélectionné; σ_i a été estimé par $\frac{2}{n_i} \left(\frac{1}{p_{E,i}} + \frac{1}{1-p_{E,i}} + \frac{1}{p_{C,i}} + \frac{1}{1-p_{C,i}} \right)$.

Si $\rho = 0$ la sélection des essais est indépendante du résultat observé; comme nous avons supposé que $\log(OR_i) < 0$ signifiait que le traitement expérimental est bénéfique, lorsque $\rho < 0$, les essais sélectionnés ont $z_i > 0$ et seront plus susceptibles d'avoir $\delta_i > 0$, donc $\varepsilon_i < 0$ et $\log(OR_i)$ surestimant $\log(\theta)$, i.e. montrant un bénéfice plus grand du traitement expérimental.

Sous ce modèle, on peut montrer que la probabilité que l'essai i soit sélectionné est

$$\tilde{p}_i = \Phi \left(\left[u_i + \tilde{\rho}_i \left((\log(OR_i) - \log(\theta)) / \sqrt{\tau^2 + \sigma_i^2} \right) \right] / \sqrt{1 - \tilde{\rho}_i^2} \right) \text{ où } u_i = a + b / s_i, \text{ et}$$

$$\tilde{\rho}_i = \sigma_i \cdot \rho / \sqrt{\tau^2 + \sigma_i^2}.$$

Nous avons fixé (a, b) en définissant des valeurs arbitraires pour les probabilités de sélectionner un essai de taille plus petite et un essai de plus grande taille: $P(\text{essai } i \text{ soit sélectionné} \mid s_i = 0,4) = 0,1$ et $P(\text{essai } i \text{ soit sélectionné} \mid s_i = 0,05) = 0,9$. On en déduit $a = -1,65$ et $b = 0,16$. Nous avons également fixé ρ . Nous avons considéré plusieurs valeurs possibles pour ρ (-0.6, -0.8, -1.0). Sachant $\log(OR_i)$ et s_i , on peut estimer \tilde{p}_i . Nous avons ensuite tiré au sort une réalisation d'une loi de Bernoulli de probabilité \tilde{p}_i . Si la valeur générée était égale à 1, l'essai était sélectionné, sinon l'essai était remplacé par un nouvel essai généré de la même manière.

Modèle de sélection en fonction de la valeur p associée à l'effet du traitement

Nous avons également utilisé l'autre classe des modèles de sélection proposés jusqu'à présent pour la modélisation du biais de publication dans la méta-analyse (Iyengar and

Greenhouse 1988; Dear and Begg 1992; Hedges and Vevea 2005). L'écriture habituelle de ces modèles de sélection implique que la densité de probabilité de l'effet du traitement observé $y_i = \log(OR_i)$ est une densité pondérée de l'effet du traitement avant sélection:

$$f(y_i | \text{trial } i \text{ selected}, s_i) = \frac{f(y_i | s_i) \cdot P(\text{trial } i \text{ selected} | y_i, s_i)}{\int_{-\infty}^{\infty} f(y_i | s_i) \cdot P(\text{trial } i \text{ selected} | y_i, s_i) dy_i}, \text{ où } f(y_i | s_i) \text{ est la fonction de}$$

densité de l'effet du traitement avant sélection (une loi normale). Si la fonction de pondération n'est pas constante, la distribution de l'effet du traitement observé diffère de celle de l'effet du traitement avant sélection, et cette différence correspond au biais de publication. Plus précisément, nous avons considéré que seuls certains essais ont été sélectionnés dans la méta-analyse selon une fonction de pondération connue $P(\text{essai } i \text{ soit sélectionné} | y_i, s_i) = w(p_i) = \exp(-\gamma_0 p_i^{\gamma_1})$ avec p_i la valeur p associée à un test bilatéral exact de Fisher pour une différence dans la proportion d'événements entre les groupes expérimental et témoin de l'essai i . Comme $w(p_i)$ est une fonction non croissante de la valeur p , les essais avec une valeur p plus petite sont plus susceptibles d'être sélectionnés que ceux avec une grande valeur p . Nous avons considéré plusieurs valeurs possibles pour (γ_0, γ_1) ($\gamma_0=4$ et $\gamma_1=3, 3/2, 3/4$). Comme nous avons fixé en γ_0 et γ_1 , $w(p_i)$ est estimable. Pour chaque essai, nous avons tiré au sort une réalisation d'une loi de Bernoulli de probabilité $w(p_i)$. Si la valeur générée était égale à 1, l'essai était sélectionné, sinon l'essai était remplacé par un nouvel essai généré de la même manière.

4.3.1.3 Simulation d'un réseau d'essais

Nous avons simulé un réseau d'essais comme J méta-analyses de $N_j = N_0$ essais chacune. Pour chaque méta-analyse j , nous avons fixé les quantités d'effet moyennes vraies θ_j . Nous avons considéré que les J quantités d'effet moyennes vraies θ_j seraient centrées autour d'un log rapport de cote ψ égal à $\psi_1 = \log(0,75)$ ou $\psi_2 = \log(0,95)$. De plus, nous avons considéré que la dispersion des quantités d'effet moyennes vraies θ_j sur les J méta-analyses

pourrait être étroite ou grande, avec une variance v du log rapport de cote égale à $v_1=0,02$ et $v_2=0,08$, respectivement. Au total, 95% des quantités d'effet moyennes vraies θ_j dans les J méta-analyses seraient distantes d'un facteur égal à $\exp(1,96 \times \sqrt{v})$ de ψ . Cela correspond à des facteurs 1,32 et 1,74 lorsque la variance vaut v_1 et v_2 , respectivement.

Par conséquent, nous avons considéré quatre scénarios en fixant des quantités d'effet moyennes vraies $\log(\theta_j)$ en tirant au sort J réalisations d'une loi normale $\log(\theta_j) \sim N(\psi, v)$ avec $\psi = \psi_1$ ou ψ_2 et $v = v_1$ ou v_2 . Les valeurs fixées sont présentées dans le tableau 4.1. Nous avons également supposé l'homogénéité de la variance inter-essais dans les J méta-analyses, c'est-à-dire $\tau_j^2 = \tau^2$. Pour une méta-analyse donnée, les données ont été ensuite générées comme précédemment.

4.3.1.4 Simulation des biais de publication dans un réseau d'essais

Nous avons induit les biais de publication sur chacune des J méta-analyses constituant le réseau comme précédemment, sous l'hypothèse que les biais étaient similaires dans les J méta-analyses. Lorsque la sélection des essais était fonction de la taille de l'essai et de l'intensité de l'effet du traitement, nous avons tiré au sort chaque ρ_j d'une distribution uniforme sur le support $[\rho_{min}, \rho_{max}]$. Lorsque la sélection des essais était fonction de la valeur p associée à l'effet du traitement, nous avons fixé γ_{0j} et nous avons tiré au sort chaque γ_{1j} d'une distribution uniforme sur le support $[\gamma_{1min}; \gamma_{1max}]$. Nous avons considéré plusieurs valeurs pour (ρ_{min}, ρ_{max}) et $(\gamma_{1min}, \gamma_{1max})$, détaillées plus bas.

Tableau 4.1 : Effets traitements moyens vrai θ_j utilisées pour générer les réseaux d'essais

Effet traitement ψ	Dispersion des effets traitements ν	Nombre de méta-analyses J	Effets traitements moyens vrais $\theta_j, j=1, \dots, J$
$\psi_1=\log(0.75)$	$\nu_1=0.02$	6	0.793, 0.889, 0.741, 0.954, 0.569, 0.684
$\psi_1=\log(0.75)$	$\nu_2=0.08$	6	0.808, 0.725, 0.876, 0.698, 0.699, 0.395
$\psi_2=\log(0.95)$	$\nu_1=0.02$	6	0.796, 1.172, 1.000, 1.171, 1.099, 0.883
$\psi_2=\log(0.95)$	$\nu_2=0.08$	6	0.491, 1.214, 0.936, 0.977, 1.451, 0.754
$\psi_1=\log(0.75)$	$\nu_1=0.02$	10	0.658, 0.852, 0.696, 0.889, 0.741, 0.722, 0.645, 0.683, 0.816, 0.796
$\psi_1=\log(0.75)$	$\nu_2=0.08$	10	0.978, 0.432, 1.149, 0.706, 0.432, 0.751, 0.679, 0.653, 0.624, 0.568
$\psi_2=\log(0.95)$	$\nu_1=0.02$	10	1.081, 1.089, 0.767, 0.973, 0.781, 0.763, 1.266, 0.816, 1.066, 0.992
$\psi_2=\log(0.95)$	$\nu_2=0.08$	10	0.845, 0.637, 1.030, 0.799, 0.541, 0.851, 1.063, 1.307, 0.674, 0.732

Nous avons fixé des quantités d'effet moyennes vraies $\log(\theta_j)$ en tirant au sort J réalisations d'une loi normale $\log(\theta_j) \sim N(\psi, \nu)$.

4.3.2 Scénarios étudiés

Dans un échantillon de 77237 essais inclus dans 14886 méta-analyses Cochrane avec données binaires, la distribution du nombre d'essais dans une méta-analyse avait une médiane égale à 3 et les percentiles à 75%, 90% et 99% valaient 6, 10 et 28, respectivement (Davey, Turner et al. 2011). Sur le même échantillon de méta-analyses, la distribution prédictive pour la variance inter-essais du log rapport de cote dans une méta-analyse future a été estimée comme étant une distribution log-normale de paramètres $(-2.56, 1.74)$. La médiane de cette distribution vaut 0,08 et les percentiles à 25% et 75% valent 0,02 et 0,25 sur l'échelle non transformée (Turner, Davey et al. 2012). Par conséquent, nous avons examiné les scénarios suivants.

Pour simuler une méta-analyse, nous avons choisi N_0 le nombre d'essais, l'effet du traitement moyen vrai θ et la variance inter-essais τ^2 . Nous avons considéré les valeurs suivantes: $N_0 \in (6, 10, 30)$; $\theta \in (0,5, 0,75, 1,0)$; $\tau^2 \in (0,02, 0,08, 0,25)$. Lorsque nous avons induit un biais de publication, nous avons généré des données jusqu'à ce que N_0 essais aient été sélectionnés avec $\rho \in (-0,6, -0,8, -1,0)$ ou avec $\beta = 4$ et $\alpha \in (3, 3/2, 3/4)$. Ces valeurs correspondent à différents degrés de biais, de modéré à sévère.

Pour simuler un réseau d'essais, nous avons choisi J le nombre de méta-analyses constituant le réseau, $N_j = N_0$ le nombre d'essais dans chaque méta-analyse, les quantités d'effet moyennes vraies θ_j et la variance inter-essais τ^2 . Nous avons considéré les valeurs suivantes : $J \in (6, 10)$, que $N_0 \in (3,6)$, θ_j comme dans le tableau 4.1, $\tau^2 \in (0,02, 0,08, 0,25)$. Lorsque nous avons induit des biais de publication, nous avons généré des données jusqu'à ce que n essais aient été sélectionnés dans chaque méta-analyse avec chaque $\rho_j \sim \text{Uniforme} [-0,8; -0,6]$ ou $\rho_j \sim \text{Uniforme} [-1,0; -0,8]$ ou avec $\beta_j = 4$ et $\alpha_j \sim \text{Uniforme} [3/2, 3]$ ou $\alpha_j \sim \text{Uniforme}$

[3/4, 3/2] pour refléter des biais similaires d'intensité modérée et sévère dans les J méta-analyses, respectivement.

4.3.3 Critères d'évaluation de la performance du test

Nous avons évalué la performance des tests dans une méta-analyse conventionnelle et dans une méta-analyse en réseau. Nous avons évalué leur taille empirique (i.e., taux d'erreur de type I) et la puissance (1 - taux d'erreur de type II) pour les scénarios avec et sans biais de publication, respectivement. Chaque statistique de test a été simulée B fois sous l'hypothèse nulle et B fois sous l'hypothèse alternative. Pour une méta-analyse en réseau, le taux d'erreur de type I était estimée par:

$$\hat{\alpha} = B^{-1} \# \{ \sum_{i=0}^N b(i|N, E/N) < 0.10 \mid \text{sans biais de publication} \}$$

et la puissance était estimée par:

$$1 - \hat{\beta} = B^{-1} \# \{ \sum_{i=0}^N b(i|N, E/N) < 0.10 \mid \text{avec biais de publication} \}.$$

En outre, comme le taux d'erreur de type I estimée peut différer de la valeur nominale, nous avons pris en compte les tailles éventuellement différentes des tests et nous avons estimé les puissances ajustées sur la taille par une analyse probit (Lloyd 2005). Considérons une statistique de test T dont les grandes valeurs conduisent au rejet de l'hypothèse nulle. Une valeur seuil pour les grandes valeurs de T est déterminée par le biais de la fonction $G(t) = \Pr(T \geq t)$. Pour un niveau d'erreur de type I nominal α^* , la valeur seuil du test est $c^* = \inf\{t: G(t) \leq \alpha^*\}$. On note la distribution vraie de T sous l'hypothèse nulle $G_0(t)$ et sous l'hypothèse non-nulle $G_I(t)$. La puissance du test peut être augmentée en diminuant la valeur seuil c^* , mais seulement au détriment de l'augmentation du niveau d'erreur de type I. Lloyd propose de représenter ce compromis par un graphe opposant la puissance $1-\beta=G_I(c^*)$ au taux d'erreur de type I $\alpha=G_0(c^*)$ lorsque la valeur seuil c^* varie. Il s'agit d'une courbe ROC que l'on peut décrire sous la forme $1-\beta=R(\alpha)=G_I(G_0^{-1}(\alpha))$. Sous l'hypothèse que la courbe ROC a une

forme probit, on approche la courbe ROC par $R(\alpha) = \Phi(\Phi^{-1}(\alpha) + \delta)$, avec δ une constante à estimer. A l'issue d'une simulation, on aura estimé $\hat{\alpha}$ sous l'hypothèse nulle et $1 - \hat{\beta}$ sous l'hypothèse non-nulle. On peut alors estimer $\hat{\delta} = \Phi^{-1}(1 - \hat{\beta}) - \Phi^{-1}(\hat{\alpha})$ puis estimer la puissance ajustée du test en calculant $\hat{R}(\alpha^*) = \Phi(\Phi^{-1}(\alpha^*) + \hat{\delta})$.

De plus, nous avons estimé le rapport de vraisemblance d'un résultat positif du test comme le rapport de la puissance (non ajustée) sur le taux d'erreur de type I, $(1 - \hat{\beta})/\hat{\alpha}$. Cela reflète la probabilité que le test donne un résultat positif sur un réseau avec biais par rapport à un réseau sans biais. Nous avons également estimé le rapport de vraisemblance d'un résultat négatif comme le rapport entre le taux d'erreur de type II (non ajustée) et 1 moins le taux d'erreur de type I, $\hat{\beta}/(1 - \hat{\alpha})$. Cela reflète la probabilité que le test donne un résultat non significatif sur un réseau avec biais par rapport à un réseau sans biais.

Enfin, dans le cas d'une méta-analyse conventionnelle, nous avons également évalué, à des fins de comparaison, la performance du test proposé par Rucker et al basé sur une régression pondérée de la transformation arc sinus des risques observés avec modélisation explicite de l'hétérogénéité entre études (Rucker, Schwarzer et al. 2008). Si nous observons x_{Ei} et x_{Ci} événements parmi n_{Ei} et n_{Ci} patients dans les groupes expérimental et témoin de l'essai i , respectivement, le test consiste à rechercher une association entre la différence des transformations arc sinus des risques d'événements $\Delta_i = \sin^{-1} \sqrt{\frac{x_{Ei}}{n_{Ei}}} - \sin^{-1} \sqrt{\frac{x_{Ci}}{n_{Ci}}}$ et sa variance approchée $I_i = \frac{1}{4 \cdot n_{Ei}} + \frac{1}{4 \cdot n_{Ci}}$. A la manière d'Egger, sur le graphique en entonnoir, on régresse Δ_i sur $\sqrt{I_i}$ en pondérant par $1/(I_i + \tau^2)$ et l'on teste si la pente associée est nulle (Egger, Davey Smith et al. 1997; Thompson and Sharp 1999).

4.3.4 Nombre de simulations à effectuer

Pour obtenir un intervalle de confiance au niveau de confiance $1 - \zeta$, pour une proportion p avec une précision pré-spécifiée A , le nombre de simulations B est $z_{1-\zeta/2}^2 \cdot p \cdot (1-p) / A^2$ avec z_{ζ} le $\zeta^{\text{ème}}$ percentile d'une loi normale centrée réduite. Pour une précision pré-spécifiée A , B atteindra son maximum pour $p = 0,5$. Pour chaque scénario, nous avons donc généré $B = 10000$ jeux de données, ce qui garantit une précision de 1% ou moins sur l'estimation de la puissance du test avec une confiance de 95%. De plus, avec ce nombre de simulations, on peut considérer que, si une estimation de l'erreur de type I vaut moins que 9,4% ou plus de 10,6%, il s'agit d'un écart significatif au niveau nominal pré-spécifié $\alpha'=0,10$.

Nous avons utilisé R v2.12.2 (R Development Core Team, Vienne, Autriche) pour toutes les analyses.

4.4 Résultats

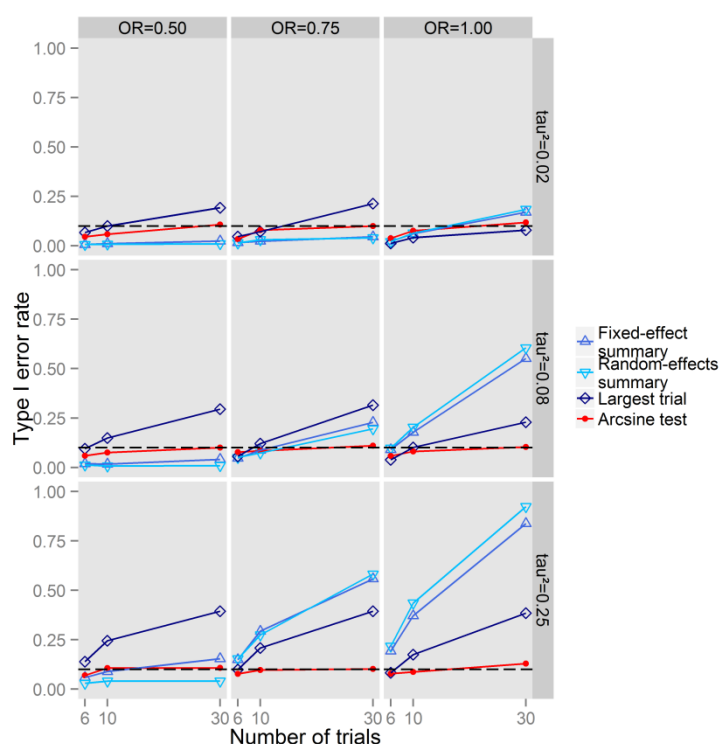
4.4.1 Méta-analyse conventionnelle

4.4.1.1 Taux d'erreur de type I

Les résultats pour les scénarios sans biais de publication sont présentés en figure 4.1. L'inspection visuelle du graphique en treillis montre que le taux d'erreur de type I augmente avec l'hétérogénéité inter-essais croissante et avec le nombre croissant d'essais dans la méta-analyse. Lorsque la quantité d'effet vraie a été estimée par l'effet combiné de la méta-analyse à effet fixe ou à effets aléatoires, le test proposé était trop conservateur avec 6 ou 10 essais, un rapport de cotes vrai de 0,5 ou 0,75, et une variance inter-essais de 0,02 ou 0,08. Par contre, le test proposé souffrait d'une inflation importante de l'erreur de type I lorsque l'hétérogénéité était importante (variance inter-essais égale à 0,25) et/ou lorsque le nombre d'essais était grand (30 essais), avec des taux d'erreur estimés supérieurs à 50%. Lorsque la quantité d'effet

vraie a été estimée par l'effet du traitement de la plus grande étude dans la méta-analyse, l'erreur de type I empirique était en accord avec le niveau de significativité pré-spécifié de 0,10, sauf lorsque le nombre d'essais était grand (30 essais), quelle que soit l'hétérogénéité, ou avec 10 essais et une variance inter-essais de 0,02 ou 0,08, auquel cas le taux d'erreur estimé variait de 14,9% à 39,4%. Enfin, le test arc sinus introduit par Rucker et al. a maintenu le mieux le taux d'erreur de type I. En effet, la taille empirique variait de 3,4% à 12,9% pour tous les scénarios. Le test était trop conservateur lorsque l'hétérogénéité et le nombre d'essais étaient faibles, sinon la taille empirique était généralement en accord avec le niveau de signification pré-spécifiée de 0,10.

Figure 4.1 : Taux d'erreur de type I des tests pour la méta-analyse conventionnelle



4.4.1.2 Puissance

Les résultats pour les scénarios avec biais de publication modélisé en fonction de la taille de l'essai et de l'intensité de l'effet du traitement estimé sont présentés en figure 4.2 ($\rho = -0,8$). L'inspection visuelle des graphiques en treillis montrent que la puissance ajustée sur

l'erreur de type I augmente avec le nombre croissant d'essais et quand l'hétérogénéité diminue. Lorsque le nombre d'essais était faible, tous les tests étaient peu puissants. Lorsque la sélection d'essais a été modélisée en fonction de la taille de l'essai et de l'intensité de l'effet du traitement, la puissance augmentait lorsque le rapport de cotes s'éloignait de 1. Le test le plus puissant était le test arc sinus. Lorsque le rapport de cotes vrai était égal à 0,5, l'hétérogénéité était faible ou modérée et le nombre d'essais était grand, la puissance du test proposé avec la quantité d'effet vraie estimée par l'effet combiné de la méta-analyse à effet fixe ou à effets aléatoires avait une puissance peu élevée mais acceptable. Dans tous les autres cas, en raison de l'inflation du taux d'erreur de type I, la puissance ajustée proposée était nettement insuffisante. Les résultats pour les autres intensités de biais étaient similaires (Annexe 2, Figures 1 et 2). Lorsque la sélection d'essais a été modélisée en fonction de la valeur p, le test le plus puissant était le test arc sinus lorsque le rapport de cotes vrai était de 0,5. Le test proposé avec la quantité d'effet vraie estimée par l'effet combiné de la méta-analyse à effet fixe ou à effets aléatoires avait une puissance acceptable lorsque le vrai rapport de cotes était égal à 1 (Annexe 2, Figures 3 à 5).

4.4.1.3 Rapports de vraisemblance

Les rapports de vraisemblance d'un résultat du test positif indiquaient que le test avait un effet modéré sur l'augmentation de la vraisemblance de biais (Figure 4.3). Les rapports de vraisemblance d'un résultat du test négatif indiquaient que le test diminuait peu la vraisemblance de biais. Le test arc sinus donnaient les meilleurs rapports de vraisemblance. Le test proposé avec la quantité d'effet vraie estimée par l'effet combiné de la méta-analyse à effets aléatoires donnait les pires rapports de vraisemblance.

Figure 4.2: Puissance des tests pour la méta-analyse conventionnelle (sélection d'essais modélisée en fonction de la taille de l'essai et de l'intensité de l'effet du traitement)

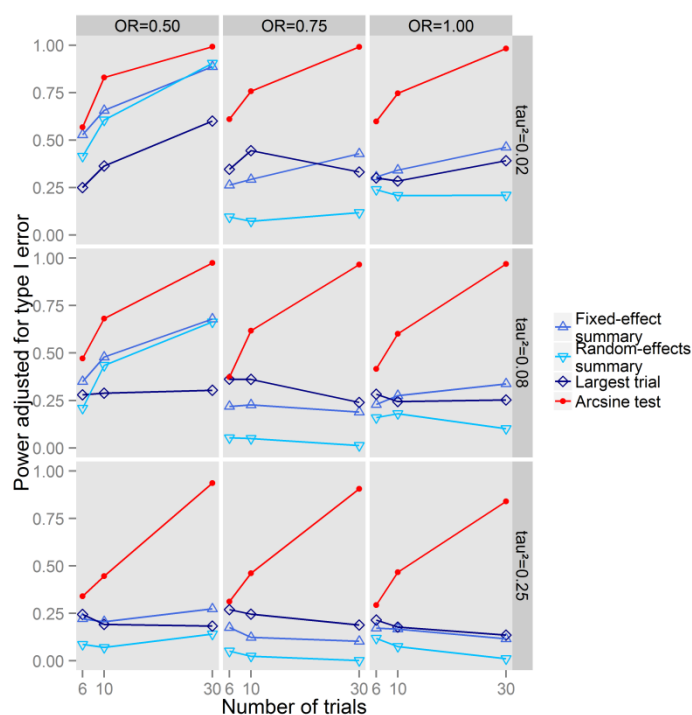
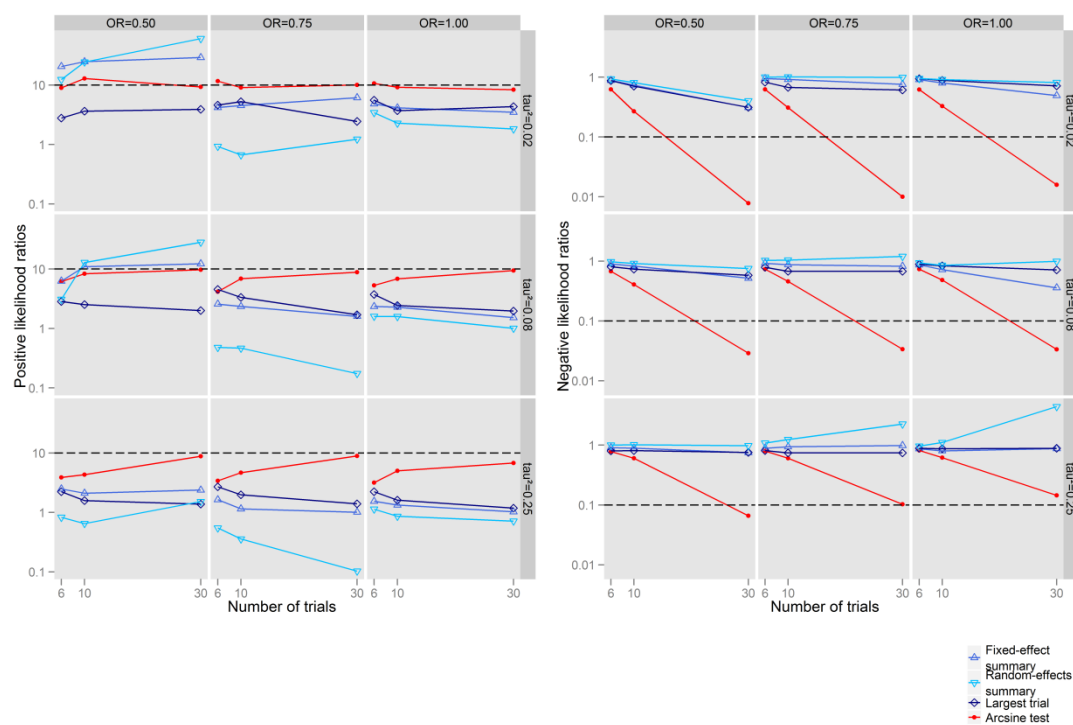


Figure 4.3: Rapports de vraisemblance des tests pour la méta-analyse conventionnelle (sélection d'essais modélisée en fonction des valeurs p associées à l'effet du traitement)

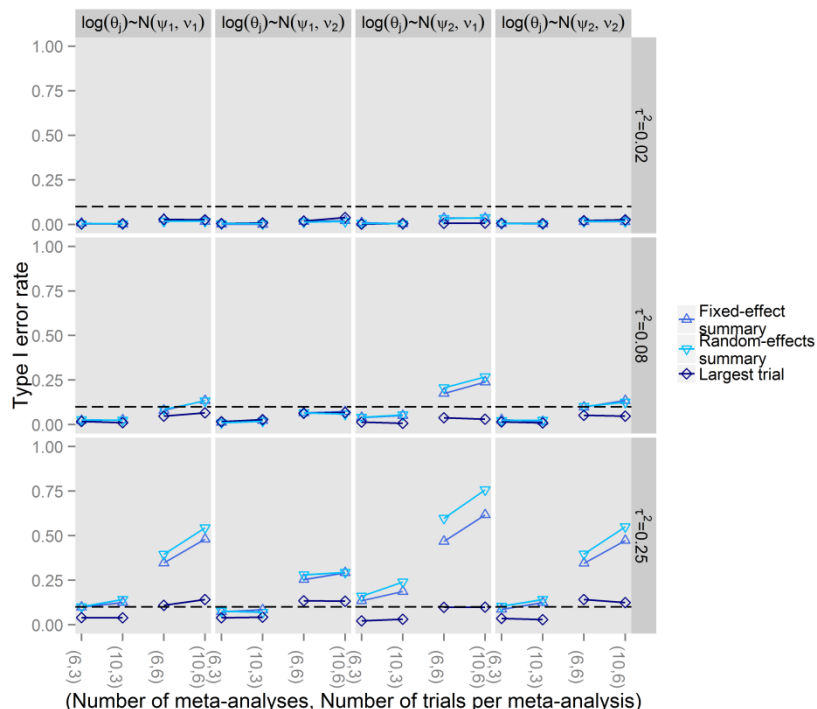


4.4.2 Méta-analyse en réseau

4.4.2.1 Taux d'erreur de type I

Les résultats pour les scénarios sans biais de publication sont présentés en figure 4.4. Le taux d'erreur de type I estimé augmentait avec l'hétérogénéité croissante et le nombre croissant de méta-analyses et d'essais par méta-analyse. Quelles que soient les quantités d'effets moyennes vraies, le test proposé était trop conservateur lorsque l'hétérogénéité intra-méta-analyse était faible ($\tau^2 = 0,02$). A l'inverse, il y avait une inflation importante du taux d'erreur lorsque l'hétérogénéité intra-méta-analyse était importante ($\tau^2 = 0,25$), excepté lorsque la quantité d'effet vraie a été estimée par l'effet du traitement de la plus grande étude dans la méta-analyse, auquel cas l'erreur estimée était en accord avec le niveau de significativité pré-spécifiée de 0,10.

Figure 4.4: Taux d'erreur de type I des tests pour la méta-analyse en réseau



4.4.2.2 Puissance

Les résultats pour les scénarios avec biais de publication modélisé en fonction de la taille de l'essai et de l'intensité de l'effet du traitement estimé sont présentés en figure 4.5 (chaque $\rho_j \sim \text{Uniforme}[-1,0; -0,8]$). Pour tous les scénarios, l'inspection visuelle des graphiques en treillis a montré que la puissance ajustée sur la taille augmentait avec le nombre croissant de méta-analyses et d'essais par méta-analyse et quand l'hétérogénéité intra-méta-analyse diminuait. Lorsqu'il y avait 6 méta-analyses et 3 essais par méta-analyse, tous les tests étaient peu puissants. Lorsque la sélection d'essais a été modélisée en fonction de la taille de l'essai et de l'intensité de l'effet du traitement, la puissance ajustée sur la taille augmentait lorsque les rapports de cotes moyens vrais s'éloignaient de 1. Le test proposé avec les quantités d'effet vraies estimées par l'effet du traitement de la plus grande étude dans chaque méta-analyse a montré une plus grande puissance ajustée que le test avec les effets combinés des méta-analyses à effet fixe et ce dernier a montré une plus grande puissance ajustée que le test avec les effets combinés des méta-analyses à effets aléatoires. Les résultats pour les autres intensités de biais étaient similaires (Annexe 2, Figure 6). Lorsque la sélection d'essais a été modélisée en fonction des valeurs p, la puissance ajustée était encore la plus petite pour le test proposé avec les effets combinés des méta-analyses à effets aléatoires, tandis que l'utilisation des effets fixes et des effets des plus grands essais donnaient des résultats relativement similaires (Annexe 2, Figures 7 et 8). Nous avons également effectué des simulations avec un autre ensemble de vecteurs d'effets du traitement moyens vrais θ_j . Les résultats étaient similaires.

4.4.2.3 Rapports de vraisemblance

Les rapports de vraisemblance d'un résultat du test positif indiquaient que le test avait un effet modéré (en cas d'hétérogénéité importante) ou important (pas ou peu d'hétérogénéité) sur l'augmentation de la vraisemblance de biais (Figure 4.6). Les rapports de

vraisemblance d'un résultat du test négatif indiquaient que le test avait un effet faible à modéré sur la diminution de la vraisemblance de biais. Les rapports de vraisemblance étaient les meilleurs lorsque les quantités d'effet vraies étaient estimées par l'effet du traitement des plus grandes études.

Figure 4.5: Puissance des tests pour la méta-analyse en réseau (sélection d'essais modélisée en fonction de la taille de l'essai et de l'intensité de l'effet du traitement)

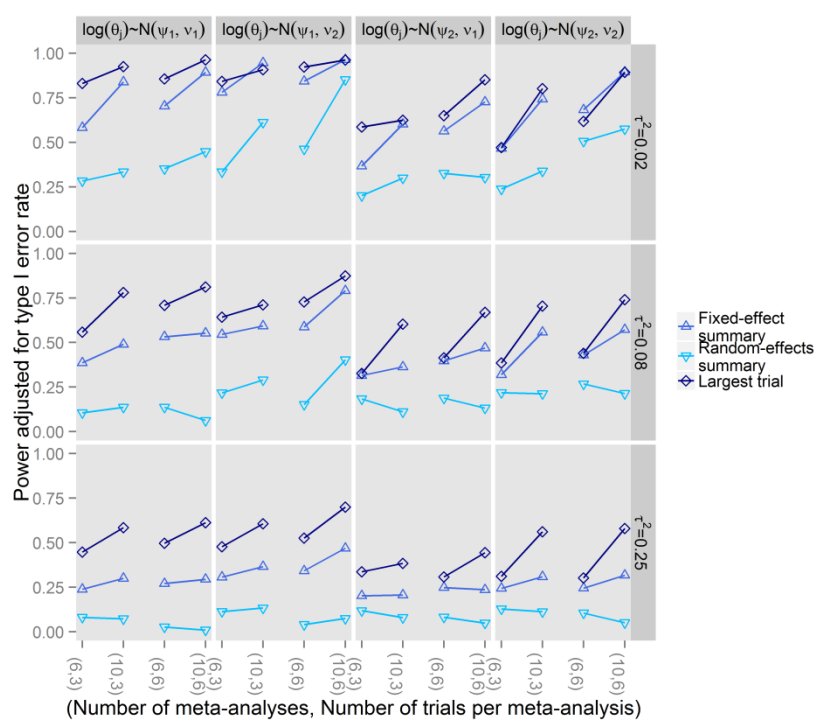
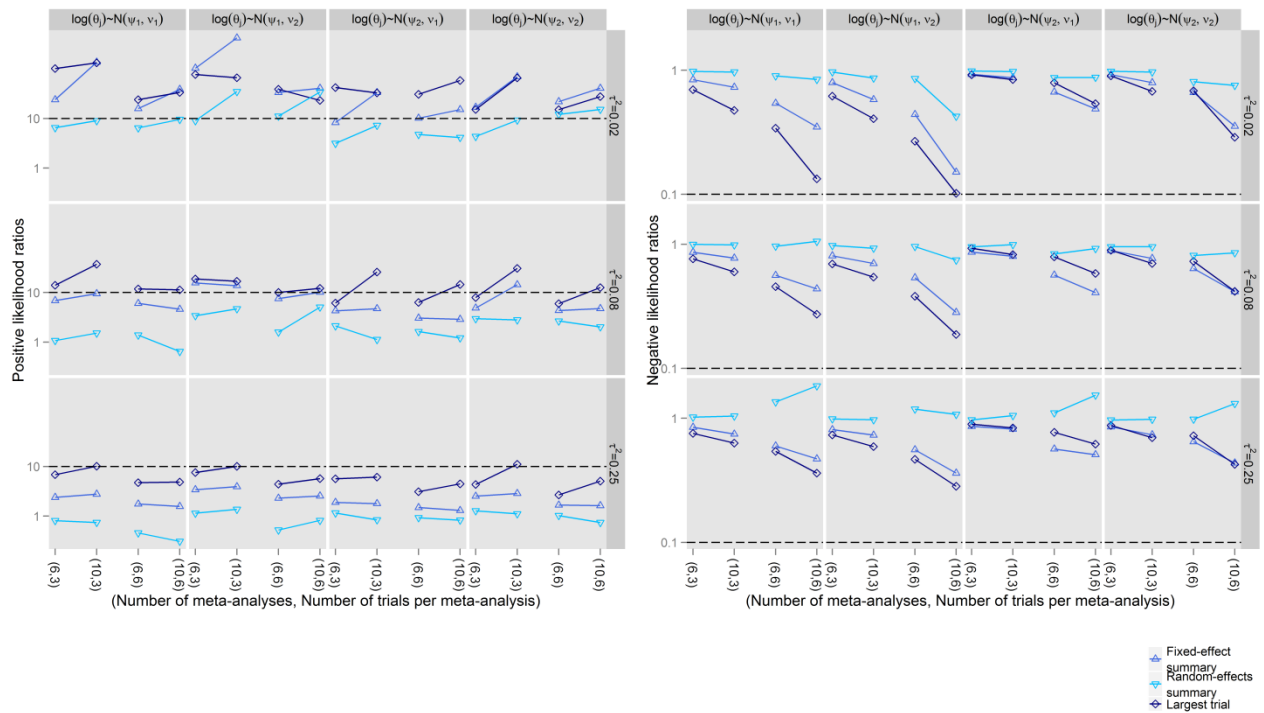


Figure 4.6: Rapports de vraisemblance des tests pour la méta-analyse en réseau (sélection d'essais modélisée en fonction des valeurs p associées à l'effet du traitement)



4.5 Application du test à deux réseaux d'essais

Nous avons appliqué le test dans deux situations où une comparaison des résultats d'essais publiés à ceux de la FDA a montré des biais de publication d'intensité élevée (réseau d'essais d'antidépresseurs) et d'intensité plus faible (réseaux d'essais d'antipsychotiques), respectivement. Les résultats issus de l'analyse des essais enregistrés à la FDA peuvent être considérés comme plus proches de l'information non biaisée.

4.5.1 Réseau d'essais d'antidépresseurs

Nous avons utilisé 2 réseaux en forme d'étoile créés à partir des rapports d'évaluation de la FDA d'essais d'antidépresseurs et de leurs publications correspondantes (Turner, Matthews et al. 2008). Turner et al ont identifié tous les essais randomisés contre placebo de 12 antidépresseurs approuvés par la FDA, puis toutes les publications correspondantes à ces essais. Les auteurs ont identifié 74 essais enregistrés à la FDA, mais ils ont constaté que 51

essais seulement avaient des résultats publiés. Les auteurs ont démontré que ces 23 essais restaient non publiés dans leur intégralité en raison de la nature négative des résultats. En outre, dans certains articles de journaux, des analyses statistiques spécifiques ont été rapportées de manière sélective et les quantités d'effet estimées différaient de celles dans les rapports de la FDA. Le critère de jugement dans cette analyse était la modification entre la randomisation et le suivi du score de sévérité des épisodes de dépression. La mesure de l'effet était une différence de moyennes standardisée. Pour illustrer le test, nous avons utilisé les quantités d'effet rapportées par Turner et al (tableau C en annexe de leur article) (Turner, Matthews et al. 2008). Parce que les résultats de deux essais évaluant la paroxétine CR ont été combinés dans l'article publié correspondant, il y avait en fait un total de 73 essais enregistrés à la FDA et 50 essais avec résultats publiés dans notre analyse.

Lorsque nous avons considéré les effets combinés par des modèles à effet fixe comme les quantités d'effet vraies, le nombre O d'essais avec résultats statistiquement significatifs observé sur le réseau de données publiées était de 41 et le nombre attendu E était de 35,3. Cet excès était statistiquement significatif au niveau pré-spécifié 0,10 (valeur p du test de probabilité binomiale = 0,05), ce qui constitue un signal pour un biais affectant les essais avec résultats publiés. Cependant, sur le réseau de données de la FDA, nous n'avons pas mis en évidence de signal d'excès de résultats statistiquement significatifs ($O=38$, $E=34.5$, $p = 0,24$). Les résultats étaient similaires lorsque nous avons utilisé les effets combinés par des méta-analyses à effets aléatoires comme effets plausibles ($O=41$, $E=35.6$, $p = 0,06$ sur données publiées; $O=38$, $E=34.6$, $p = 0,25$ pour les données FDA). Par contre, lorsque nous avons utilisé les effets estimés par les essais les plus grands comme effets plausibles, il y avait un signal d'excès de résultats statistiquement significatifs encore plus fort sur les données publiées et il y avait également un signal positif sur le réseau de données de la FDA ($O=41$, $E=31.3$, $p = 0,002$ sur données publiées; $O=38$, $E=29.0$, $p = 0,02$ pour les données FDA). Ce

dernier n'est pas nécessairement un faux positif : même si les données de la FDA sont plus complètes que les données publiées, il est possible qu'il existe encore certains biais dans ces données (Ioannidis 2008).

4.5.2 Réseau d'essais d'antipsychotiques

Turner et al. ont également utilisé les données de la FDA pour évaluer si l'efficacité apparente des antipsychotiques de seconde génération avait été influencée par des biais de publication (Turner, Knoopfmlacher et al. 2012). Les auteurs ont identifié les essais randomisés contre placebo de phase 2/3 pour huit antipsychotiques approuvés par la FDA, puis tous les articles publiés correspondant aux rapports d'évaluation. Les auteurs ont identifié 24 essais enregistrés par la FDA parmi lesquels 20 ont vu leurs résultats publiés. Le critère de jugement était l'évolution de la randomisation au suivi de la sévérité des symptômes de schizophrénie. La mesure d'effet traitement était la différence de moyennes standardisée. Nous avons utilisé les estimations rapportées par Turner et al, Tableaux S2 et S3 (Turner, Knoopfmlacher et al. 2012).

Lorsque nous avons considéré les effets combinés par des modèles à effet fixe comme les quantités d'effet vraies, nous n'avons pas mis en évidence de signal d'excès de résultats statistiquement significatifs sur le réseau de données publiées ($O=19$, $E=18.1$, $p = 0,43$) ni sur le réseau de données FDA ($O=20$, $E=19.6$, $p = 0,53$). Les résultats étaient similaires lorsque nous avons utilisé les effets combinés par des méta-analyses à effets aléatoires comme effets plausibles ($O=19$, $E=18.4$, $p = 0,50$ sur données publiées; $O=20$, $E=19.7$, $p = 0,56$ pour les données FDA). Par contre, lorsque nous avons utilisé les effets estimés par les essais les plus grands comme effets plausibles, il y avait un signal d'excès de résultats statistiquement significatifs sur les données publiées ($O=19$, $E=16.1$, $p = 0,097$) mais pas sur le réseau de données de la FDA ($O=20$, $E=18.6$, $p = 0,36$).

4.6 Discussion

Dans cet article, nous avons proposé un test pour la détection du biais de publication dans la méta-analyse en réseau, quel que soit le type de critère de jugement (données binaires, continues ou de survie). Le test est basé sur les nombres observé et attendu d'essais avec résultats statistiquement significatifs sur le réseau. A l'issue d'études de simulation, nous avons montré que le taux d'erreur de type I était en accord avec le niveau nominal et que le test était raisonnablement puissant après ajustement sur l'erreur de type I, sauf lorsque l'hétérogénéité inter-essais était importante auquel cas le taux d'erreur de type I était considérablement augmenté et la puissance faible.

Le test correspond bien à l'idée largement répandue selon laquelle une estimation de l'effet du traitement doit franchir un seuil de significativité statistique, ce qui conduit à une aversion pour les résultats « négatifs », ou à l'inverse, à une poursuite du résultat significatif (Ioannidis 2005; Fanelli 2012). En outre, le test ne fait aucune hypothèse sur l'effet des petites études. Enfin, un excès de résultats significatifs pourrait refléter la suppression des résultats de certains essais en intégralité, mais aussi la présentation sélective de certaines analyses statistiques ou de certains critères de jugement, ou même la fabrication de données même si cette situation est plus rare (Fanelli 2009). En effet, les biais de publication englobent plusieurs mécanismes qui peuvent opérer séparément ou simultanément, mais chacun d'entre eux peut être motivé par la recherche de la significativité statistique.

Dans une méta-analyse en réseau, les tests d'asymétrie conventionnels pourraient être appliqués à chaque méta-analyse constituant le réseau. Si le biais de publication est détecté dans une des comparaisons directes, les méta-analystes ont un signal pour interpréter les résultats de la synthèse avec beaucoup de prudence. Toutefois, le nombre d'essais dans chaque comparaison directe est souvent limité (moins de 10 essais pour chaque comparaison directe), ce qui empêcherait cette approche de documenter ou d'exclure le biais de publication de

manière appropriée (Ioannidis and Trikalinos 2007; Sterne, Sutton et al. 2011). De plus, tester la présence de biais sur chaque méta-analyse conventionnelle dans le réseau l'une après l'autre pourrait donner à des faux-positifs. Ici, le test proposé tire sa force de l'ensemble des essais à travers le réseau sous l'hypothèse que les biais sont interchangeable, c'est-à-dire que les biais, s'ils sont présents, opèrent d'une manière similaire sur tous les essais du réseau. Comme nous l'avons discuté ailleurs, si nous n'avons pas d'information pour distinguer différents mécanismes de biais sur différentes branches du réseau, l'interchangeabilité des biais est plausible (Trinquart, Chatellier et al. 2012).

Les décisions sur le caractère concluant et la diffusion des résultats sont généralement basées (uniquement) sur la significativité statistique, et cette pratique est susceptible d'affecter toutes les branches d'un réseau. Nous reconnaissons que l'hypothèse ne peut pas être tenable dans des contextes où les biais de publication peuvent affecter le réseau de façon déséquilibrée. Par exemple, seuls certaines comparaisons directes pourraient être affectées et pas les autres. Cependant, le test proposé pourrait encore être utile dans de tels réseaux s'il y a un nombre suffisant d'essais dans les comparaisons directes affectées. Enfin, les méta-analystes devraient considérer que, comme dans les essais randomisés par grappes (en anglais, cluster randomized trials), une plus grande puissance peut être obtenue en augmentant le nombre de branches dans le réseau, par rapport à l'augmentation du nombre d'essais sur chaque branche. Ceci est en faveur de l'élargissement d'un réseau, par exemple en étendant l'ensemble des interventions éligibles et les comparateurs, afin de faciliter l'exploration des biais.

Pour estimer le nombre d'essais avec des résultats statistiquement significatifs, l'estimation de la quantité d'effet vraie pour chaque méta-analyse dans le réseau est nécessaire. Premièrement, nous notons que, en présence de biais de publication, l'effet combiné issu d'une méta-analyse à effet fixe ou à effets aléatoires est susceptible d'être biaisé

et de surestimer l'effet vrai. Même l'effet estimé par la plus grande étude dans la méta-analyse pourrait être parfois biaisé, et souvent, il n'y aura pas d'essai suffisamment grand. Par conséquent, les effets plausibles qui sont utilisés sont conservateurs pour tester l'excès de résultats statistiquement significatifs. Deuxièmement, comme une méta-analyse à effets aléatoires donne plus de poids aux petits essais relativement à une méta-analyse à effet fixe, l'effet combiné issue de la méta-analyse à effets aléatoires peut être le plus éloigné de l'effet vrai en présence de biais de publication. Cela peut expliquer que la puissance du test proposé est moins bonne lorsque l'on utilise les effets combinés par effets aléatoires comme effets plausibles. Enfin, nous avons exploré l'utilisation des essais les plus précis et nous avons constaté que le test proposé avait une puissance satisfaisante. Par conséquent, nous recommandons, pour les applications pratiques sur des réseaux d'essais, l'utilisation de l'effet fourni par l'essai le plus grand ou l'estimation par méta-analyse à effet fixe. La première méthode peut avoir un petit avantage sur la deuxième. La transformation arc sinus est aussi un excellent choix vu ses performances sur une méta-analyse conventionnelle, mais cette métrique reste peu utilisée dans la littérature et ne peut pas être utilisée pour combiner les résultats dans une méta-analyse en réseau. Cependant, par extension, il est possible d'appliquer la transformation arc sinus à toutes les méta-analyses conventionnelles d'un réseau dans le but de tester l'excès de résultats significatifs.

Un problème potentiel avec le test proposé est que le biais de publication et l'hétérogénéité pourraient être confondus. En effet, le taux d'erreur de type I augmentait avec la variance inter-essais croissante. C'est un problème typique de tous les tests de biais de publication introduits pour la méta-analyse conventionnelle. Cependant, nous avons observé ce résultat lorsque la variance inter-essais est égale à 0,25 (25% seulement des méta-analyses ont ce degré d'hétérogénéité) et en supposant que toutes les méta-analyses constituant le réseau avaient ce degré d'hétérogénéité, ce qui est vraisemblablement peu fréquent en

pratique. Le cas échéant, on devrait sans doute reconsidérer l'ensemble de l'information disponible avant de réaliser une synthèse ; les sources d'hétérogénéité devraient être explorées et le test pourrait éventuellement être appliqué à un sous-ensemble d'essais plus homogènes. Enfin, il convient de noter que les biais de publication peuvent induire eux-mêmes de l'hétérogénéité (Ioannidis and Trikalinos 2007; Jackson 2007). Quoi qu'il en soit, le test doit être utilisé et interprété à bon escient. Dans le cas d'un signal de biais statistiquement significatif, il ne faut pas conclure que ce biais est un biais de publication spécifiquement, il pourrait refléter d'autres pratiques telles que la présentation sélective des analyses ou des critères de jugement. En l'absence de signal de biais significatif, il ne faut pas exclure la possibilité d'un biais de publication, parce qu'un résultat négatif du test ne diminue pas considérablement la probabilité a priori de biais, sauf s'il y a beaucoup d'essais ou si l'hétérogénéité inter-essais est importante (Ioannidis 2008).

Bien que nous ayons illustré le test sur des réseaux en étoile, le test peut être utilisé sur des réseaux avec boucles fermées. A ce propos, un avantage de notre méthode de simulation est que les effet moyens vrais générés peuvent refléter des réseaux d'essais contre placebo, des réseaux d'essais face-face ou des réseaux avec à la fois des essais contre placebo et des essais face-face. Néanmoins, nos études de simulation présentent des limites. Tout d'abord, les valeurs de plusieurs variables ont été dérivées à partir d'un large échantillon de méta-analyses Cochrane afin d'avoir des scénarios réalistes. Cependant, on ne peut exclure que certaines méta-analyses de cet échantillon aient été affectées par des biais, notamment des biais de publication. Or ces biais de publication pourraient avoir influencé les nombres d'essais observés, la distribution des tailles d'essais et de l'hétérogénéité (Jackson 2007). Deuxièmement, nous avons constaté que l'erreur de type I et la puissance du test proposé étaient variables en fonction de caractéristiques du réseau connues par les méta-analystes et en fonction d'autres qui elles seront toujours inconnues, notamment la nature du mécanisme de

sélection des essais. Nous avons abordé ce problème en générant des données selon deux modèles de sélection, utilisés précédemment dans la littérature, l'un selon la taille des essais et l'intensité des effets estimés (Copas and Shi 2000; Copas and Shi 2001; Rucker, Schwarzer et al. 2008), l'autre en fonction de la valeur p associée à l'effet du traitement (Begg and Mazumdar 1994; Macaskill, Walter et al. 2001; Harbord, Egger et al. 2006), et nous avons trouvé des résultats similaires. Une limite supplémentaire est que nous avons considéré que les risques d'événements vrais étaient entre 0,3 et 0,7. Le test perdrait sans doute de la puissance si les risques d'événements étaient plus faibles ; cependant, aucun test n'a de bonne performance dans de telles situations. Si quelques méta-analyses dans le réseau ont des risques d'événements dans l'intervalle 0,3-0,7, ce sont elles qui contribueront principalement à la puissance du test.

D'autres approches de modélisation ont été introduites récemment pour évaluer la présence de biais de publication dans une méta-analyse en réseau. Mavridis et al. ont présenté une mise en œuvre bayésienne du modèle de sélection Copas étendu à une méta-analyse en réseau (Mavridis, Sutton et al. 2012). Le paramètre de corrélation ρ y est supposé égal pour toutes les comparaisons directes et indique le degré de biais de publication. Des modèles de méta-régression en réseau qui permettent à la quantité d'effet de dépendre de l'écart type ou de la variance associée ont été également décrits (Chaimani and Salanti 2012; Trinquart, Chatellier et al. 2012). Sous l'hypothèse de biais interchangeable, une pente de régression moyenne a posteriori est estimée et indique l'intensité de l'effet des petites études sur le réseau.

4.7 Conclusion

Le test proposé pourrait être utile pour fournir un signal statistique pour un excès de résultats significatifs dans la méta-analyse en réseau. Si le biais de publication est détecté sur

le réseau ou dans toute comparaison directe en utilisant des approches conventionnelles, la méta-analyse en réseau doit être interprétée avec prudence.

Chapitre 5

Correction du biais de publication dans la méta-analyse en réseau

5.1 Contexte

5.2 Méthodes

5.3 Résultats

5.4 Discussion

5.5 Conclusion

5.1 Contexte

Les méta-analyses en réseau sont de plus en plus utilisées pour évaluer la meilleure intervention parmi différentes interventions thérapeutiques disponibles pour une maladie donnée. Le principe de base de cette approche est que l'intervention A est comparée à un comparateur C, puis l'intervention B avec C et une comparaison indirecte ajustée permet de comparer A et B, malgré l'absence de tout essai randomisé face-face comparant A et B. Une méta-analyse en réseau généralise ce principe et permet de résumer les données comparatives pour de multiples interventions en combinant des comparaisons directes et indirectes (Lumley 2002; Lu and Ades 2004; Salanti, Higgins et al. 2008). L'objectif est d'estimer les quantités d'effet pour toutes les comparaisons deux à deux possibles entre interventions, bien qu'aucun essai ne soit disponible pour certaines comparaisons.

Le biais de publication est une menace majeure pour la validité des résultats des revues systématiques ou méta-analyses conventionnelles (Dwan, Altman et al. 2008; Song, Parekh et al. 2010). La prise en compte des biais de publication dans la méta-analyse en réseau est difficile, car la disponibilité inégale des résultats de l'ensemble du réseau peut compromettre la validité de la synthèse (Salanti, Kavvoura et al. 2008; Li, Puhan et al. 2011).

Nous avons déjà évalué empiriquement l'impact des biais de publication sur les résultats de la méta-analyse en réseau d'essais d'antidépresseurs et nous avons montré que ces biais peuvent biaiser les estimations de l'efficacité du traitement (Trinquart, Abbé et al. 2012).

De nombreuses méthodes ont été proposées comme analyses de sensibilité pour évaluer la robustesse des méta-analyses conventionnelles au biais de publication et aux biais liés à l'effet des études de petite taille (Hedges 1992; Silliman 1997; Larose and Dey 1998; Copas and Shi 2000; Duval and Tweedie 2000; Sutton, Song et al. 2000; Copas and Shi 2001; Preston, Ashby et al. 2004; Bowden, Jackson et al. 2010; Carpenter, Rucker et al. 2011; Rucker, Carpenter et al. 2011; Rufibach 2011). Ces méthodes de modélisation incluent des approches basées sur la régression et des modèles de sélection. Nous étendons ces deux approches à la méta-analyse en réseau dans le cadre bayésien.

5.2 Méthodes

Premièrement, nous avons étendu un modèle de méta-régression, reliant la quantité d'effet à l'écart type, récemment décrit pour la méta-analyse conventionnelle (Moreno, Sutton et al. 2009; Moreno, Sutton et al. 2012). Dans cette approche, la pente de régression reflète l'intensité de l'association entre la quantité d'effet et la précision (i.e., l'intensité de l'effet des petites études; en anglais, *small-study effect*), et la constante donne une quantité d'effet combinée et ajustée (la quantité d'effet prédite pour un essai de précision infinie). Deuxièmement, nous avons introduit un modèle de sélection, qui modélise la probabilité qu'un essai d'être observé. Cette dernière est prise en compte par pondération inverse dans la méta-analyse en réseau. Les deux modèles d'ajustement reposent sur l'hypothèse que les biais sont interchangeables sur le réseau, c'est-à-dire que les biais, s'ils existent, opèrent de manière similaire sur l'ensemble des essais du réseau. Troisièmement, nous avons appliqué ces modèles d'ajustement à des jeux de données créés à partir des rapports d'évaluation par la

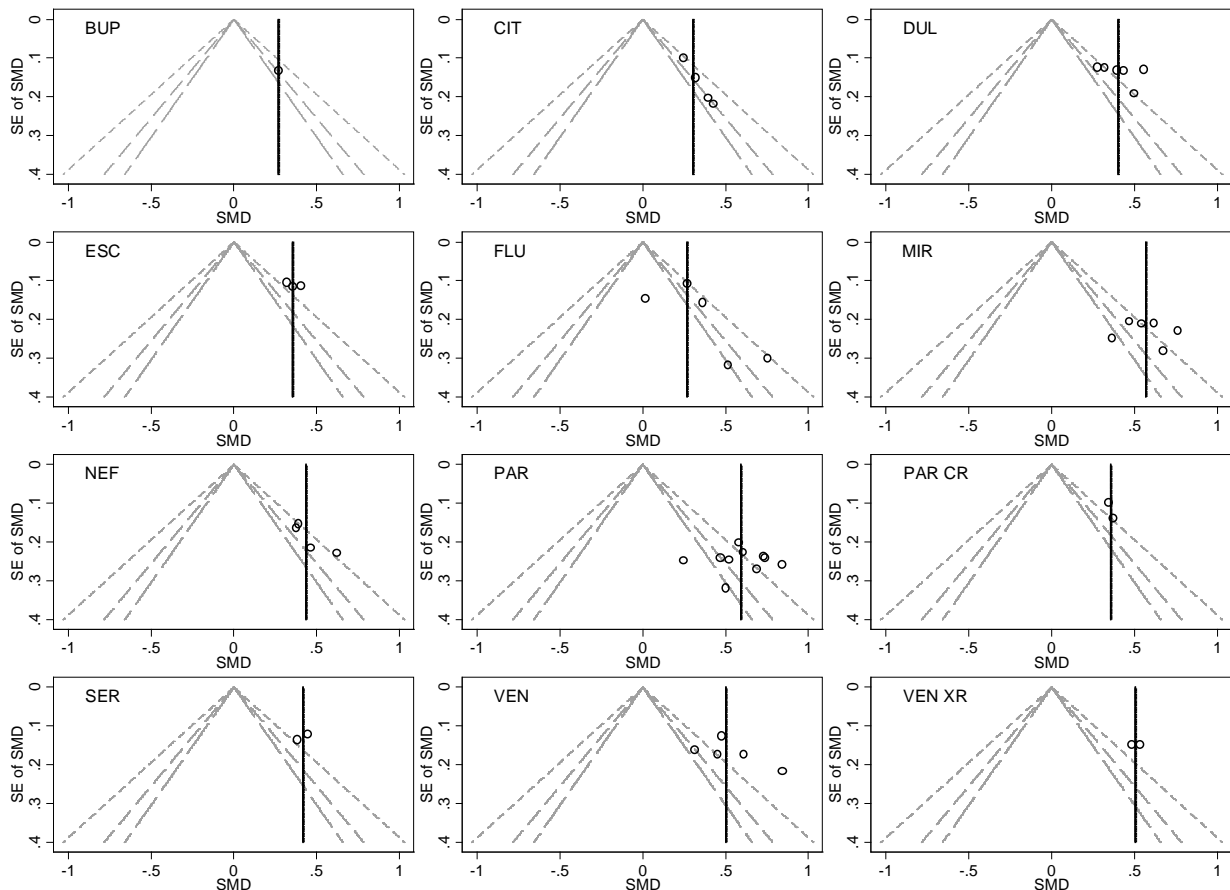
FDA d'essais d'antidépresseurs et à partir de leurs publications correspondantes. Il a déjà été démontré que ces jeux de données différaient en raison de biais de publication (Turner, Matthews et al. 2008). Nous avons comparé les résultats obtenus via les modèles d'ajustement appliqués aux données publiées à ceux obtenus via le modèle de méta-analyse en réseau standard appliqué aux données publiées puis aux données de la FDA, ces dernières étant considérées comme la norme de référence.

5.2.1 Bases de données utilisées

Turner et al. ont évalué dans une revue antérieure la publication sélective des essais d'antidépresseurs (Turner, Matthews et al. 2008). Les auteurs ont identifié tous les essais contrôlés randomisés versus placebo de 12 antidépresseurs approuvés par la FDA. Puis ils ont identifié les publications correspondantes à ces essais en cherchant les bases de données de la littérature et en contactant les promoteurs d'essais. A partir de la base de données de la FDA, les auteurs ont identifié 74 essais. Les résultats de 23 essais dans leur intégralité n'ont pas été publiés. La proportion d'essais dont les résultats n'ont pas été publiés variait selon les médicaments, allant de 0% pour la fluoxétine et la paroxétine CR à 60% et 67% pour la sertraline et le bupropion (Annexe 3, Tableau 1). Les résultats de ces essais n'étaient pas publiés à cause de la nature (négative) des résultats. Par ailleurs, dans certains articles publiés, des analyses statistiques spécifiques ont été rapportées de manière sélective et les quantités d'effet rapportées étaient différentes de celles rapportées par la FDA. Dans cette analyse, le critère de jugement était la modification de la randomisation au suivi du score de sévérité des épisodes de dépression. La mesure de l'effet traitement était une différence moyenne standardisée. Des méta-analyses conventionnelles des données de la FDA ont montré une efficacité diminuée pour tous les médicaments par rapport aux données publiées, la diminution de l'effet allant de 10% et 11% pour la fluoxétine et la paroxétine CR à 39% et

41% pour la mirtazapine et la nefazodone (Annexe 3, Tableau 1). La figure 5.1 montre les graphiques en entonnoir pour les données publiées. L'inspection visuelle ne suggère pas d'effet traitement plus fort dans les essais de petite taille (i.e., pas d'asymétrie du graphique en entonnoir) sur aucune des 12 comparaisons entre médicament et placebo.

Figure 5.1 Graphique en entonnoir pour les essais d'antidépresseurs avec résultats publiés



Chaque graphique en entonnoir est le nuage de points des estimations de l'effet du traitement par chaque essai contre les écarts types associés, la ligne verticale pleine représentant l'effet traitement combiné. En l'absence de biais de publication, on pourrait s'attendre à un graphique en entonnoir (en anglais, funnel plot) symétrique. Il arrive que le graphique en entonnoir ne soit pas symétrique, ce qui peut être dû à un biais de publication, mais il existe d'autres sources possibles d'asymétrie. Les lignes de niveaux en pointillés représentent les seuils de significativité statistique (tiret long $p=0,1$; tiret $p=0,05$; tiret court $p=0,01$). Si les études semblent faire défaut dans les zones de non-significativité, l'asymétrie peut être due au biais de publication plutôt qu'à d'autres facteurs.

5.2.2 Méta-analyse en réseau

Le modèle standard pour la méta-analyse en réseau a été formalisé par Lu et Ades (Higgins and Whitehead 1996; Lu and Ades 2004; Lu and Ades 2006). Nous supposons que

chaque essai i a évalué les traitements j et k parmi T interventions dans le réseau. Chaque essai a fourni une estimation de l'effet traitement y_{ijk} de j relativement à k et sa variance associée v_{ijk} . Nous supposons que $y_{ijk} > 0$ indique la supériorité de j relativement à k . En faisant l'hypothèse d'une vraisemblance normale et selon un modèle à effets aléatoires, $y_{ijk} \sim N(\theta_{ijk}, v_{ijk})$ et $\theta_{ijk} \sim N(\theta_{ijk}, \tau^2)$, où θ_{ijk} est l'effet vrai sous-jacent à chaque comparaison randomisée entre les traitements j et k et θ_{ijk} est l'effet moyen vrai de j relativement à k . Le modèle suppose que la variance inter-essais est homogène (i.e., $\tau_{jk}^2 = \tau^2$). Cette hypothèse peut être relaxée (Lu and Ades 2004; Schöttker, Luhmann et al. 2009). Le modèle suppose également la cohérence entre l'information directe et indirecte: si l'on considère le traitement b comme le traitement de référence de l'ensemble du réseau, les effets des traitements j, k , etc. par rapport au traitement b sont considérés comme des paramètres de base, et les contrastes restants, appelés paramètres fonctionnels, sont déduits des équations de cohérence $\theta_{jk} = \theta_{jb} - \theta_{kb}$ pour chaque $j, k \neq b$.

5.2.3 Modèles d'ajustement

5.2.3.1 Modèle de méta-régression

Nous avons utilisé un modèle de méta-régression en réseau, en généralisant une approche basée sur la régression pour ajuster la méta-analyse conventionnelle sur l'effet des études de petite taille (Stanley 2008; Moreno, Sutton et al. 2009; Moreno, Sutton et al. 2009; Rucker, Schwarzer et al. 2011; Moreno, Sutton et al. 2012). Cette approche basée sur la régression tient compte d'un éventuel effet des études de petite taille en reliant la quantité d'effet à une mesure de sa précision. Ici, nous supposons une relation linéaire entre la quantité d'effet et son écart type, et le modèle implique une extrapolation au-delà des données observées à une étude hypothétique de précision infinie. Ce modèle s'écrit:

$$y_{ijk} \sim N(\gamma_{ijk}, v_{ijk})$$

$$\gamma_{ijk} = \theta_{ijk} + \mathbb{I}_{ijk} \cdot \beta_{jk} \cdot \sqrt{v_{ijk}}$$

$$\beta_{jk} \sim N(\beta, \sigma^2)$$

$$\theta_{ijk} \sim N(\theta_{jk}, \tau^2)$$

$$\theta_{jk} = \theta_{jb} - \theta_{kb} \text{ pour tout } j, k \neq b$$

La figure 1.A en Annexe 3 montre une représentation graphique du modèle. Dans l'équation de régression, θ_{ijk} est l'effet du traitement ajusté sur l'effet des études de petite taille sous-jacent à chaque comparaison randomisée entre les traitements j et k ; β_{jk} représente l'effet potentiel des études de petite taille (i.e., la pente associée à l'asymétrie du graphique en entonnoir pour les comparaisons randomisées entre les traitements j et k). Le modèle suppose que ces pentes de régression spécifiques à chaque comparaison entre traitements suivent une distribution normale commune, avec une pente moyenne β et une variance σ^2 . Ceci équivaut à supposer que les biais liés aux études de petite taille spécifiques à chaque comparaison sont en fait échangeables au sein du réseau. Puisque nous avons supposé que $y_{ijk} > 0$ indique la supériorité de j relativement à k , $\beta > 0$ signifie qu'il existe une tendance générale pour un effet des petites études (i.e., les contrastes entre traitements ont tendance à être surestimés dans les petits essais). Enfin, \mathbb{I}_{ijk} est une indicatrice égale à 1 si l'on s'attend à ce que l'effet des petites études favorise le traitement j relativement à k , égale à -1 dans le cas contraire, et égale à 0 lorsque l'on n'a aucune raison de croire qu'il existe un biais dans un sens ou dans l'autre (par exemple, pour un essai face-face comparant deux traitements actifs aussi récents l'un que l'autre). Dans les études comparant des traitements actif et inactif (par exemple, placebo ou absence d'intervention), on peut raisonnablement s'attendre à ce que le traitement actif soit toujours favorisé par le biais lié aux petits essais.

5.2.3.2 Modèle de sélection

Nous utilisons un modèle qui tient compte du biais de publication en utilisant une fonction de pondération pour représenter les processus de sélection. Le modèle comprend un modèle de la quantité d'effet (c'est-à-dire, le modèle standard de méta-analyse en réseau qui spécifie les distributions des estimations des effets traitement en l'absence de mécanisme de sélection), et un modèle de sélection qui spécifie comment les distributions des effets traitement sont modifiées par les processus de sélection (Song, Eastwood et al. 2000; Sutton, Song et al. 2000). Nous supposons que la probabilité de sélection dépend de l'écart type de la quantité d'effet, comme une fonction décroissante de celui-ci. Nous adoptons une approche basée sur un modèle de sélection logistique, comme précédemment utilisé dans la méta-analyse classique (Hedges 1988; Carpenter, Rucker et al. 2011).

$$y_{ijk} \sim N(\gamma_{ijk}, v_{ijk})$$

$$\gamma_{ijk} = \theta_{ijk} / w_i$$

$$\text{logit } w_i = \beta_{0jk} + \beta_{1jk} \cdot \mathbb{I}_{ijk} \cdot \sqrt{v_{ijk}}$$

$$\beta_{0jk} \sim N(\beta_0, \sigma_0^2) \text{ et } \beta_{1jk} \sim N(\beta_1, \sigma_1^2)$$

$$\theta_{ijk} \sim N(\theta_{jk}, \tau^2)$$

$$\theta_{jk} = \theta_{jb} - \theta_{kb} \text{ pour tout } j, k \neq b$$

La figure 1.B en Annexe 3 montre une représentation graphique du modèle. Dans l'équation de régression logistique, w_i représente la propension d'un essai à être publié, β_{0jk} définit la probabilité globale d'observer une comparaison randomisée entre les traitements j et k et β_{1jk} contrôle la vitesse à laquelle cette probabilité évolue lorsque l'écart type associé à l'effet traitement estimé augmente. Nous nous attendons à ce que β_{1jk} soit négatif de sorte que les résultats des essais produisant des écarts types plus grands aient moins tendance à être publiés. Le modèle suppose l'interchangeabilité des coefficients β_{0jk} et β_{1jk} au sein du réseau. En fixant $\gamma_{ijk} = \theta_{ijk} / w_i$, nous définissons un modèle simple qui surpondère les

résultats d'essais avec propension plus faible d'être publiés afin qu'ils aient une influence disproportionnée dans le modèle de méta-analyse en réseau. θ_{ijk} est l'effet traitement ajusté sur le processus de sélection qui sous-tend chaque comparaison randomisée entre les traitements j et k . Enfin, \mathbb{I}_{ijk} est défini de la même manière que dans la section précédente.

5.2.4 Estimation des modèles

Nous avons estimé 4 modèles: le modèle standard de méta-analyse en réseau sur les données publiées, les deux modèles d'ajustement sur les données publiées, et le modèle standard de méta-analyse en réseau sur les données de la FDA. Dans chaque cas, le modèle a été estimé via la méthode de Monte Carlo par chaînes de Markov (algorithme d'échantillonnage de Gibbs). Le placebo a été naturellement choisi comme le traitement de référence global pour comparer tous les autres traitements. Par conséquent, les 12 effets entre chaque agent antidépresseur et placebo sont les paramètres de base. Pour 2 traitements j et k , $SMD_{jk} > 0$ indique que j est supérieur à k . Dans les deux modèles de méta-régression et de sélection, nous avons supposé que les traitements actifs seraient toujours favorisés par les biais liés aux petites études par rapport au placebo; par conséquent, \mathbb{I}_{ijk} est toujours égal à 1.

Dans le modèle standard de méta-analyse en réseau, nous avons défini des distributions a priori pour les paramètres de base θ_{jb} et la variance commune τ^2 : $\theta_{jb} \sim N(0, 100^2)$ et $\tau \sim \text{Uniforme}(0, 10)$. Dans le modèle de méta-régression, nous avons choisi des lois a priori vagues pour la pente moyenne β et entre la variance inter-pentes commune σ^2 : $\beta \sim N(0, 100^2)$ et $\sigma \sim \text{Uniforme}(0, 10)$. Dans le modèle de sélection, nous avons choisi des lois a priori faiblement informatives pour les mesures de tendance centrale et de dispersion (β_0, σ_0^2) et (β_1, σ_1^2). Nous avons considérés p_{min} et p_{max} la probabilité qu'un essai soit publié lorsque l'écart type prend ses valeurs minimale et maximale sur le réseau de données publiées et nous avons spécifié des lois a priori Beta pour ces probabilités (Bedrick,

Christensen et al. 1996). Nous l'avons réalisé indirectement en spécifiant des valeurs attendues pour la médiane et le 5^{ème} ou 95^{ème} percentile (Wu, Shih et al. 2008). Pour les essais avec écart type égal à la valeur minimale observée, nous avons supposé que les chances que p_{min} soit $< 50\%$ étaient de 5% et les chances que $p_{min} < 80\%$ étaient de 50%. Pour les essais avec écart type égal à la valeur maximale observée, notre hypothèse était que les chances que p_{max} soit $< 40\%$ étaient de 50% et les chances que $p_{max} < 70\%$ étaient de 95%. Nous discutons de ces choix plus loin dans la discussion. A partir de cette information, nous avons déterminé que les distributions a priori pour p_{min} et p_{max} étaient les lois Beta (7,52; 2,63) et Beta (3,56; 4,84), respectivement. Enfin, nous avons exprimé β_0 et β_1 en termes de p_{min} et p_{max} . Pour les écarts types σ_0 et σ_1 , nous avons choisi des distributions uniformes sur (0,2). Pour chaque analyse, nous avons construit des distributions a posteriori à partir de 2 chaînes de 500000 simulations, après que la convergence ait été atteinte sur 500000 simulations initiales pour chaque chaîne (en anglais, burn-in). Nous avons utilisé WinBUGS v1.4.3 (Imperial College et du CRM, Londres, Royaume-Uni) pour estimer tous les modèles bayésiens, et R v2.12.2 (R Development Core Team, Vienne, Autriche) pour résumer les résultats et évaluer la convergence. Les codes sont rapportés en Annexe 3, Texte 1.

5.2.5 Comparaison des modèles

Nous avons comparé les résultats des 2 modèles d'ajustement appliqués aux données publiées et les résultats du modèle standard de méta-analyse en réseau appliqué aux données publiées et aux données FDA, ce dernier cas étant la norme de référence. Premièrement, nous avons comparé les moyennes a posteriori et les intervalles de crédibilité à 95% pour les 12 paramètres de base et la variance commune, ainsi que pour les 66 paramètres fonctionnels (i.e., $12 \times 11/2 = 66$ comparaisons deux à deux possibles entre les 12 médicaments). Deuxièmement, nous avons comparé les classements des traitements. Nous avons évalué la

probabilité que chaque traitement soit le meilleur, puis le deuxième meilleur, le troisième meilleur, etc. Nous avons tracé les probabilités cumulées et calculé la surface sous la courbe des probabilités de rang cumulées (SUCRA) pour chaque traitement (Salanti, Ades et al. 2010). Troisièmement, afin de comparer les différents modèles appliqués aux données publiées, nous avons utilisé la moyenne a posteriori de la déviance résiduelle et le critère d'information de déviance (Spiegelhalter, Best et al. 2002).

5.3 Résultats

Dans le modèle de méta-régression en réseau appliqué aux données publiées, la pente moyenne a posteriori β était de 1,7 (intervalle de crédibilité à 95% -0,3 à 3,6), ce qui suggère une tendance générale pour un effet des études de petite taille dans le réseau. Les 12 pentes de régression étaient similaires, avec des moyennes postérieures allant de 1,4 à 1,9. Dans le modèle de sélection appliqué aux données publiées, la pente moyenne β_I était de -10,0 (-18,0 à 2,50), indiquant que la propension à être publié avait tendance à être plus faible pour les essais avec écart type plus grand. Dans les deux modèles, toutes les estimations étaient soumises à une incertitude importante (Annexe 3, Tableau 2).

Le tableau 5.1 présente les estimations des 12 paramètres de base entre chaque médicament et le placebo selon les 4 modèles. Par rapport à la méta-analyse en réseau des données publiées, les deux modèles d'ajustement sur données publiées ont montré que l'ensemble des 12 effets estimés des médicaments par rapport au placebo a été réduit. Pour le modèle de méta-régression, la diminution de l'efficacité variait de 48% pour la venlafaxine XR à 99% pour la fluoxétine. Pour le modèle de sélection, la diminution variait de 13% pour l'escitalopram à 26% pour la paroxétine. Lorsque l'on considère les paramètres fonctionnels (i.e., les 66 comparaisons deux à deux possibles entre les 12 médicaments), nous avons constaté des différences entre les résultats des modèles d'ajustement et ceux du modèle

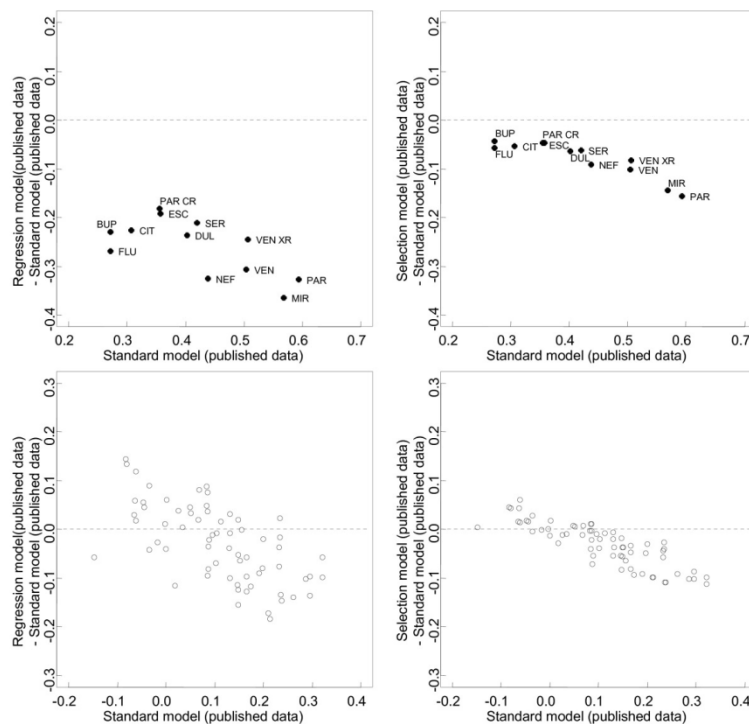
standard de méta-analyse en réseau appliqué aux essais publiés (Figure 5.2). La différence médiane relative, en valeur absolue, entre les quantités d'effet deux à deux via le modèle de méta-régression en réseau et le modèle standard de méta-analyse en réseau était de 57,3% (25% - 75% percentile 30,3% - 97,6%), la différence médiane relative entre le modèle de sélection et le modèle standard était de 29,2% (15,1% - 46,1%).

Tableau 5.1 Comparaison des estimations issues des 2 modèles d'ajustement appliqués aux données publiées à celles issues du modèle de MA en réseau standard appliqué aux données publiées et aux données de la FDA

	Données de la FDA		Données publiées	
	Modèle standard	Modèle de régression	Modèle de sélection	Modèle standard
	Moyenne (Ecart type)	Moyenne (Ecart type)	Moyenne (Ecart type)	Moyenne (Ecart type)
θ_{BUP}	0.176 (0.081)	0.043 (0.256)	0.229 (0.121)	0.271 (0.139)
θ_{CIT}	0.240 (0.074)	0.081 (0.171)	0.254 (0.073)	0.306 (0.076)
θ_{DUL}	0.300 (0.054)	0.166 (0.190)	0.340 (0.066)	0.402 (0.058)
θ_{ESC}	0.310 (0.067)	0.165 (0.193)	0.311 (0.070)	0.357 (0.068)
θ_{FLU}	0.256 (0.081)	0.004 (0.160)	0.215 (0.068)	0.271 (0.074)
θ_{MIR}	0.351 (0.070)	0.206 (0.331)	0.424 (0.110)	0.567 (0.092)
θ_{NEF}	0.256 (0.076)	0.112 (0.260)	0.348 (0.094)	0.437 (0.094)
θ_{PAR}	0.426 (0.063)	0.267 (0.346)	0.438 (0.105)	0.593 (0.078)
$\theta_{PAR\ CR}$	0.323 (0.101)	0.174 (0.187)	0.309 (0.083)	0.354 (0.085)
θ_{SER}	0.252 (0.077)	0.210 (0.231)	0.359 (0.094)	0.419 (0.094)
θ_{VEN}	0.395 (0.071)	0.199 (0.224)	0.403 (0.092)	0.504 (0.075)
$\theta_{VEN\ XR}$	0.398 (0.094)	0.261 (0.273)	0.423 (0.110)	0.506 (0.107)
τ	0.060 (0.037)	0.031 (0.024)	0.024 (0.019)	0.032 (0.025)

Les données sont les moyennes a posteriori et les écarts types des paramètres de base (θ), l'hétérogénéité entre les essais (τ)

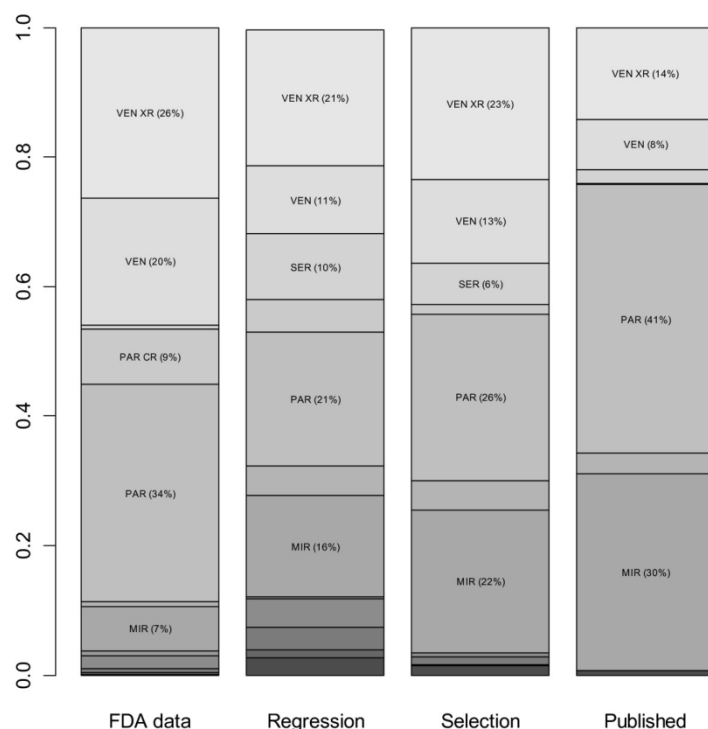
Figure 5.2 Différence entre les estimations des comparaisons deux à deux entre les 12 agents antidépresseurs et le placebo: modèle de régression sur données publiées par rapport à la MA en réseau standard sur données publiées (panneau de gauche); modèle de sélection sur données publiées par rapport à la MA en réseau standard sur données publiées (panneau de droite).



Les abscisses montrent les estimations du modèle standard appliqué aux données publiées, les axes des ordonnées montrent les différences entre les estimations issues des modèles d'ajustement (régression ou sélection) sur données publiées et les estimations issues du modèle standard sur données publiées. Les points noirs sont les 12 effets des médicaments par rapport au placebo; les points blancs sont les 66 comparaisons deux à deux possibles entre les 12 médicaments.

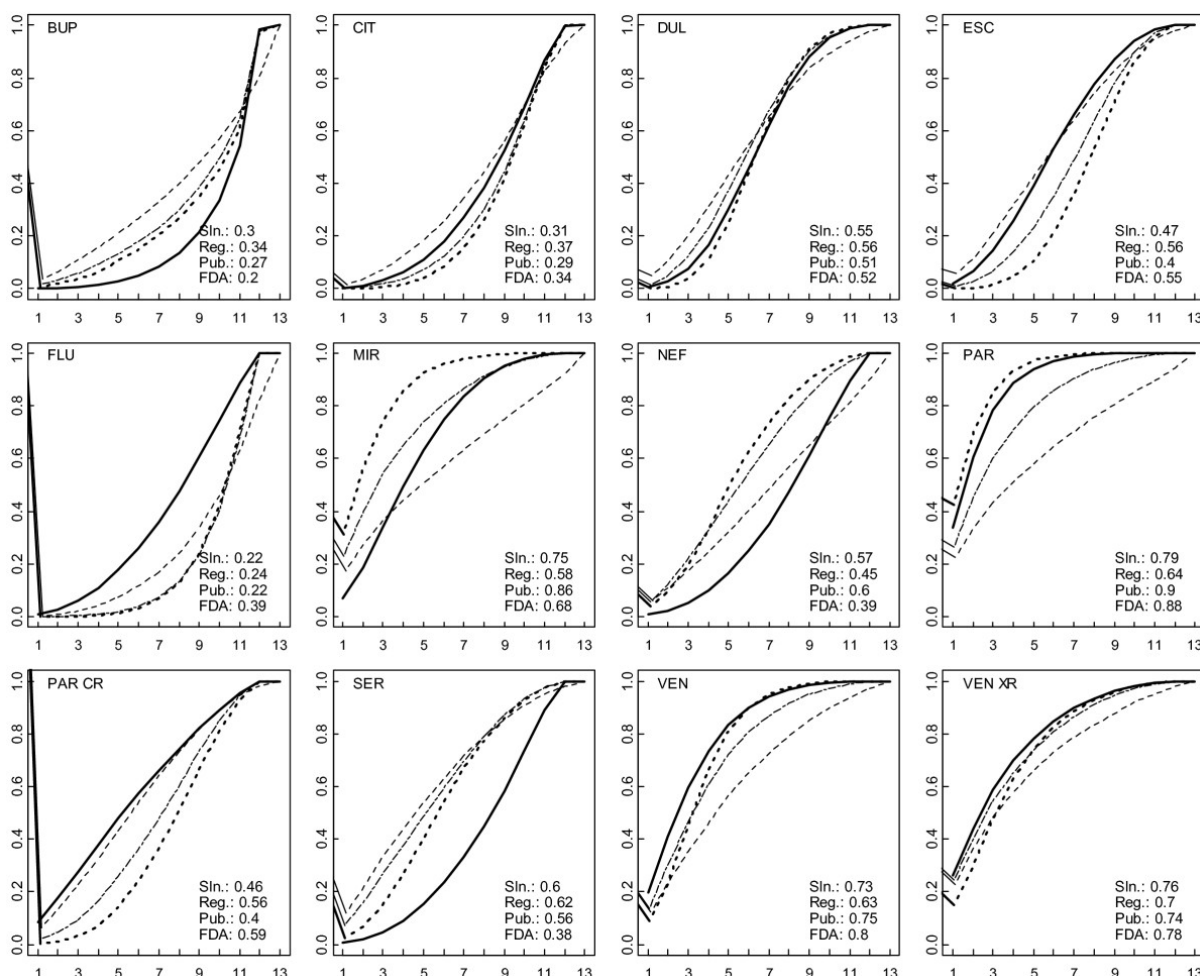
La figure 5.3 résume les probabilités d'être le meilleur antidépresseur. Par rapport à la méta-analyse en réseau standard des données publiées, les probabilités d'être le meilleur antidépresseur étaient réduites avec les modèles d'ajustement sur données publiées pour la paroxétine (de 41,5% à 20,7% avec le modèle de régression ou 25,7% avec le modèle de sélection) et la mirtazapine (de 30,3% à 15,7% ou 21,9%). A l'inverse, cette probabilité était augmentée pour la venlafaxine (de 7,9% à 10,6% ou 12,8%) et la venlafaxine XR (de 14,1% à 21,0% ou 23,5%).

Figure 5.3 : Probabilités que chaque antidépresseur soit le meilleur selon la méta-analyse en réseau standard des données de la FDA, le modèle de méta-régression, de sélection ou la méta-analyse en réseau standard des données publiées



La figure 5.4 montre les courbes des probabilités de rang cumulatives et les aires sous la courbe (SUCRA). Pour la méta-analyse en réseau standard des données publiées, la paroxétine et la mirtazapine sont à égalité à la première place et la venlafaxine XR et la venlafaxine sont à égalité au troisième rang. Le modèle de sélection appliqué aux données publiées a donné un classement légèrement différent, la paroxétine, la mirtazapine et la venlafaxine XR à égalité à la première place et la venlafaxine en quatrième place. Pour le modèle de méta-régression en réseau appliqué aux données publiées, la venlafaxine XR était au premier rang, la venlafaxine et la paroxétine à égalité au deuxième rang et la mirtazapine était cinquième.

Figure 5.4: Courbes des probabilités de rang cumulées pour les 12 antidépresseurs : modèle standard appliqué aux données de la FDA (ligne solide en gras) et aux données publiées (ligne en pointillés gras) et modèles d'ajustement appliqués aux données publiées, modèle de régression (ligne en pointillés) et modèle de sélection (ligne en double pointillés).



Sur chaque graphique, l'axe des abscisses représente les rangs possibles de $r = 1$ à $r = 13$ et l'axe des ordonnées représente les probabilités cumulées que le traitement correspondant soit parmi les r meilleurs traitements. Plus la courbe est dans le coin supérieur gauche, plus le traitement est efficace. La surface sous la courbe vaut 1 lorsqu'un traitement est le meilleur et 0 quand un traitement est le pire. FDA: modèle standard appliqué aux données FDA; Pub.: modèle standard appliqué aux données publiées; Reg.: modèle de régression appliquée aux données publiées; Sln.: modèle de sélection appliqué aux données publiées.

Pour les modèles d'ajustement appliqués aux données publiées, l'hétérogénéité inter-essais et la qualité de l'ajustement étaient comparables à celles obtenues avec la méta-analyse en réseau standard des données publiées (Tableaux 5.1 et 5.2).

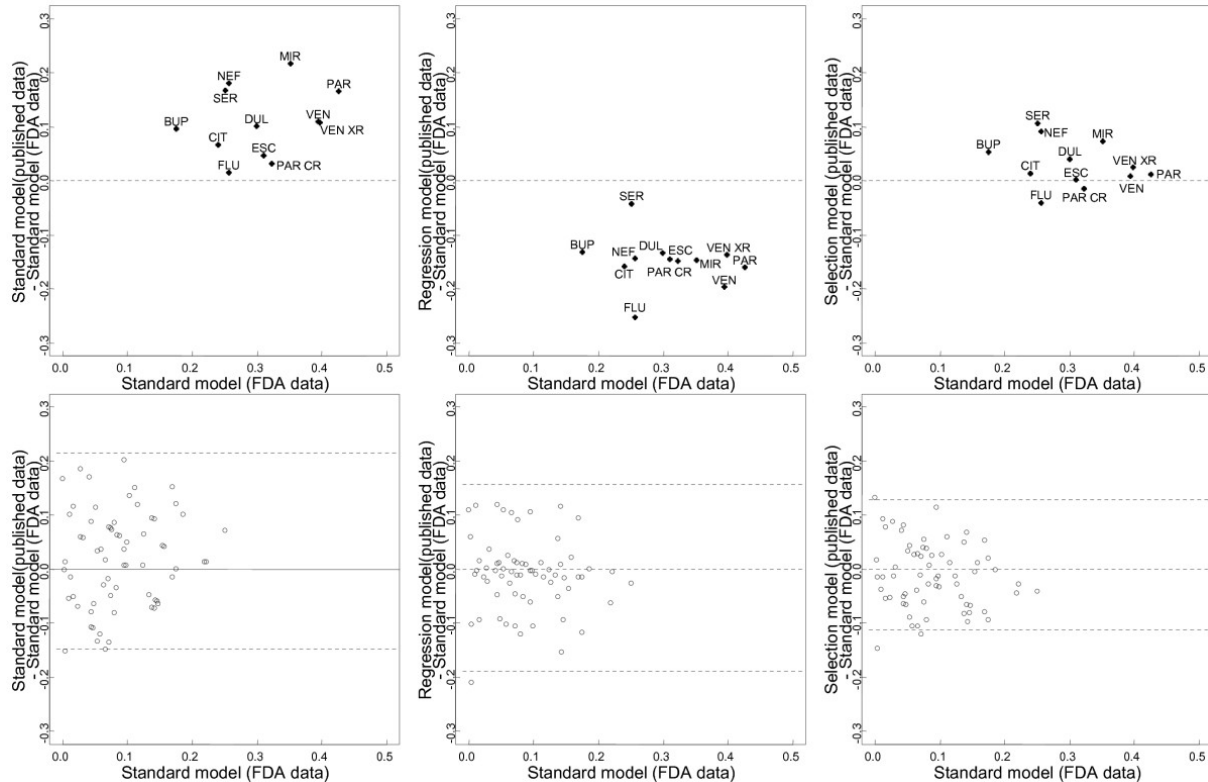
Tableau 5.2: Comparaison de l'ajustement et de la complexité entre les 2 modèles d'ajustement et le modèle standard, tous appliqués aux données publiées

	Modèle de Régression	Modèle de Sélection	Modèle standard
Déviance résiduelle (\bar{D}_{res})	31.4	31.5	34.4
Nombre effectif de paramètres (pD)	15.9	14.7	13.9
Critère de déviance (Deviance Information Criteria)	47.3	46.2	48.3

Des valeurs plus faibles de \bar{D}_{res} indiquent un meilleur ajustement aux données. Des valeurs plus faibles de DIC indiquent un meilleur compromis entre l'ajustement du modèle et la complexité du modèle. Une différence de DICs de 5 ou plus peut être considérée comme substantielle (<http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/dicpage.shtml>)

Les effets des médicaments estimés par rapport au placebo dans les modèles de régression et de sélection sont similaires à ceux de la méta-analyse en réseau des données de la FDA pour certains médicaments (Tableau 5.1). Il y avait des différences lorsque l'on considère les 66 comparaisons deux à deux possibles entre les médicaments (Figure 5.5). Les résultats différaient également entre modèles pour la probabilité d'être le meilleur médicament et le classement des médicaments. Dans la méta-analyse en réseau des données de la FDA, la probabilité d'être le meilleur médicament était de 7,3% pour la mirtazapine, 33,9% pour la paroxétine, 19,3% pour la venlafaxine et de 25,7% pour la venlafaxine XR; la paroxétine était au premier rang, et la venlafaxine et la venlafaxine XR à égalité au deuxième rang (Figure 5.3).

Figure 5.5: Différence entre les estimations des comparaisons deux à deux entre les 12 agents antidépresseurs et le placebo: méta-analyse en réseau standard sur données publiées par rapport à la méta-analyse en réseau standard sur données FDA (à gauche); modèle de régression sur données publiées par rapport à la méta-analyse en réseau standard sur données FDA (au milieu); modèle de sélection sur données publiées par rapport à la méta-analyse en réseau standard sur données FDA (à droite).



Les abscisses montrent les estimations du modèle standard NMA appliquée aux données de la FDA, les axes des ordonnées montrent les différences entre les estimations des modèles d'ajustement (régression ou sélection) sur données publiées et les estimations du modèle standard sur données de la FDA. Les points noirs sont les 12 effets des médicaments estimés par rapport au placebo; les points blancs sont les 66 comparaisons possibles deux à deux entre les 12 médicaments.

5.4 Discussion

Nous avons généralisé deux méthodes d'ajustement au biais de publication de la méta-analyse à la méta-analyse en réseau. La première méthode combine méta-analyse en réseau et modèle de méta-régression, et les quantités d'effet sont régressées sur leur précision. La seconde méthode combine le modèle de méta-analyse en réseau avec un modèle de sélection logistique qui permet d'estimer la probabilité qu'un essai soit publié ou sélectionné dans le réseau. La première méthode ajuste essentiellement sur l'asymétrie du graphique en entonnoir (en anglais, funnel plot) ou, autrement dit, sur les effets des petites études, qui peuvent

résulter d'autres causes que le biais de publication. La seconde méthode ajuste sur le biais de publication (i.e., la suppression d'un essai dans son ensemble en fonction des résultats). Les deux modèles tirent partie de l'ensemble des essais dans le réseau, avec l'hypothèse que les biais opèrent d'une manière similaire sur tous les essais dans le domaine.

Dans un réseau spécifique d'essais d'antidépresseurs, randomisés contrôlés contre placebo, basé sur des données décrites et publiées précédemment par Turner et al., la comparaison des résultats obtenus via les modèles d'ajustement appliqués aux données publiées à ceux obtenus via le modèle de méta-analyse en réseau standard appliqué aux données publiées a permis d'évaluer la robustesse des estimations d'efficacité relative et du classement au biais de publication ou biais liés aux petites études. Les deux modèles ont montré une diminution de toutes les valeurs estimées des paramètres de base (i.e., les 12 quantités d'effet pour chaque agent par rapport au placebo). Les 66 contrastes deux à deux entre médicaments, les probabilités d'être le meilleur médicament et le classement ont été également modifiés. La méta-analyse en réseau des données publiées n'était pas robuste au biais de publication et aux biais liés aux petites études.

Ce jeu de données spécifique offrait la possibilité d'effectuer une méta-analyse en réseau à la fois sur des données publiées et sur des données de la FDA. Ces dernières peuvent être considérées comme une « source de données sans biais (bien qu'incomplète) » pour les essais d'antidépresseurs randomisés contre placebo (Moreno, Sutton et al. 2009). La comparaison des résultats des 2 modèles d'ajustement appliqués aux données publiées à ceux du modèle de méta-analyse en réseau standard appliqué aux données de la FDA a montré que les quantités d'effet entre agent antidépresseur et placebo ont été corrigées pour certains médicaments mais pas tous. Ceci a conduit à des différences entre modèles d'ajustement et analyse standard des données FDA dans les 66 contrastes deux à deux entre les médicaments, les probabilités d'être le meilleur médicament et le classement des agents. Par conséquent, les

2 modèles ne devraient pas être considérés comme optimaux; l'objectif n'est pas de produire des estimations définitives corrigées des biais de publication et autres biais liés aux petites études, mais plutôt d'évaluer la robustesse des résultats à un biais potentiel.

Des approches similaires ont été utilisées par d'autres auteurs. Des modèles de méta-régression en réseau dans un cadre bayésien ont déjà été utilisés pour évaluer l'impact des biais liés à l'innovation (en anglais, novelty bias) et des défauts méthodologiques dans les essais (Dias, Welton et al. 2010; Salanti, Dias et al. 2010). Un modèle de méta-régression en réseau pour évaluer spécifiquement l'impact de l'effet des études de petite taille a été utilisé par Dias et al. dans une ré-analyse d'un réseau d'essais randomisés face-face d'inhibiteurs sélectifs du recaptage de la sérotonine (Dias, Welton et al. 2010). En suivant l'approche basée sur la régression de Moreno et al. en méta-analyse classique, les auteurs ont introduit une mesure de la taille de l'étude comme une variable dans le modèle et ont identifié un biais moyen sur les contrastes deux à deux entre traitements. Plus récemment, Moreno et al. ont utilisé une approche similaire pour corriger l'effet des études de petite taille dans plusieurs méta-analyses conventionnelles évaluant des interventions similaires sur des critères de jugement similaires et ont illustré leur méthode en utilisant le jeu de données de Turner et al. (Moreno, Sutton et al. 2011). Notre approche diffère de la leur en ce que nous avons étendu cette approche par méta-régression pour la méta-analyse en réseau. Nous avons utilisé l'écart type de l'estimation de l'effet traitement comme variable de régression. De plus, nous avons spécifié une variance inter-essais additive plutôt qu'un paramètre de surdispersion multiplicatif. En effet, ce paramètre multiplicatif peut être estimé inférieur à 1, ce qui implique moins d'hétérogénéité qu'attendu à cause des fluctuations d'échantillonnage. D'autres approches par modèle de sélection ont été considérées récemment. Chootrakool et al. ont introduit un modèle avec approximation normale basé sur les log rapports de cotes empiriques pour les méta-analyses en réseau dans un cadre fréquentiste et ils ont appliqué des modèles de

sélection de Copas à certains groupes d'essais dans le réseau sélectionnés en fonction de l'asymétrie sur les graphiques en entonnoir (Chootrakool, Shi et al. 2011). Mavridis et al. ont présenté une mise en œuvre bayésienne du modèle de sélection de Copas étendu aux méta-analyses en réseau et ont appliqué leur méthode sur le réseau de Turner et al. (Mavridis, Sutton et al. 2012). Dans le modèle de sélection de Copas, la probabilité de sélection dépend à la fois des estimations des effets traitement et de leurs écarts types. Dans l'extension à la méta-analyse en réseau, un coefficient de corrélation ρ supplémentaire, supposé égal pour toutes les comparaisons, doit être estimé. Lorsqu'ils sont appliqués aux données publiées du réseau de Turner et al., le modèle de sélection que nous avons proposé et le modèle de sélection de Mavridis et al. donnent des résultats similaires.

Les 2 modèles d'ajustement reposent sur l'hypothèse d'interchangeabilité des processus de sélection à travers le réseau, c'est-à-dire que les biais, s'ils sont présents, opèrent de manière similaire sur les essais du réseau. Dans cette étude de cas, par construction, toutes les études étaient des essais randomisés contre placebo, enregistrés auprès de la FDA, promus et financés par l'industrie pharmaceutique, et pour tous les médicaments, certains résultats négatifs n'étaient pas publiés (Turner, Matthews et al. 2008). Ainsi, l'hypothèse d'interchangeabilité des processus de sélection est plausible. Plus généralement, si nous n'avons pas d'information pour distinguer différents mécanismes de biais de publication sur le réseau, une distribution a priori interchangeable est plausible, « l'ignorance impliquant l'interchangeabilité » (Gelman, Carlin et al. 2004; McCandless, Gustafson et al. 2012). Toutefois, l'hypothèse pourrait ne pas être tenable dans d'autres contextes dans lesquels les biais de publication peuvent affecter le réseau de façon déséquilibrée. Ces biais peuvent fonctionner différemment dans les essais randomisés contre placebo et les essais face-face (Rising, Bacchetti et al. 2008), dans des essais plus anciens et plus récents (à cause de registres d'essais), et pour les interventions médicamenteuses et non médicamenteuses (Li,

Puhan et al. 2011). Dans les réseaux plus complexes impliquant des essais face-face, les 2 modèles d'ajustement pourraient être généralisés pour permettre au biais de publication ou au biais des petites études attendus pour les essais face-face de différer de ceux attendus dans les essais comparant des traitements actifs et inactifs (Dias, Welton et al. 2010). Dans les essais face-face, la direction attendue du biais est incertaine, mais on pourrait supposer que le traitement sponsorisé est favorisé (biais de financement; en anglais, sponsoring bias) (Lexchin, Bero et al. 2003; Lathyris, Patsopoulos et al. 2010) ou que le traitement le plus récent est favorisé (biais d'optimisme) (Chalmers and Matthews 2006; Bero, Oostvogel et al. 2007; Salanti, Dias et al. 2010). Si j est le traitement produit par l'industriel pharmaceutique qui a sponsorisé l'essai et k le traitement comparateur, \mathbb{I}_{ijk} serait égal à 1. Ou \mathbb{I}_{ijk} serait égal à 1 si le traitement j est plus récent que le traitement k . Cependant, il serait encore difficile de distinguer les sources de biais opérant sur l'information directe et l'information indirecte, surtout si les biais de publication et l'incohérence sont présents simultanément ou si les directions attendues des biais sont en conflit sur une boucle.

Les modèles que nous avons décrits ont des limites. Premièrement, ils conduiraient à de mauvaises estimations des quantités d'effets et des biais lorsque les méta-analyses conventionnelles au sein du réseau comprennent un petit nombre d'essais (Moreno, Sutton et al. 2009). Deuxièmement, pour le modèle de sélection, nous avons spécifié la fonction de pondération. Si les hypothèses sous-jacentes (i.e., une forme de lien logistique et la probabilité d'observer un essai fonction de l'écart type associé à l'effet traitement estimé) sont erronées, le modèle de sélection estimé sera erroné. Toutefois, d'autres fonctions de pondérations (par exemple, via le lien probit) ou d'autres méthodes de conditionnement (par exemple, sur l'intensité de la quantité d'effet) pourraient être envisagées. Enfin, le modèle de sélection a été mis en œuvre avec une loi a priori peu informative, qui spécifiait que la propension qu'une étude soit publiée puisse diminuer avec l'augmentation de l'écart type. Il existe un risque que

cette information a priori écrase le poids des données observées, en particulier si le nombre d'essais est faible. Bien qu'elles aient été quelque peu arbitraires, ces lois a priori pour les paramètres du modèle de sélection sont en accord avec les valeurs fixées dans les études précédentes utilisant le modèle de sélection de Copas (Copas and Shi 2000; Rucker, Schwarzer et al. 2008). Plusieurs mécanismes de biais de sélection pourraient être testés, par exemple, en faisant varier les valeurs modales a priori pour p_{min} et p_{max} , les probabilités de publication lorsque l'écart type prend ses valeurs minimales et maximales sur le réseau (Copas and Shi 2001).

5.5 Conclusion

En conclusion, nous avons montré que la prise en compte des biais de publication et biais liés aux petites études était possible dans cette étude de cas. La validité des méthodes introduites est conditionnée par un nombre suffisant d'essais dans le réseau et par l'hypothèse que les méta-analyses conventionnelles constituant le réseau partagent un biais moyen commun. Des analyses de simulation sont nécessaires pour déterminer dans quelles conditions ces modèles d'ajustement sont valides. L'application des modèles d'ajustement devrait être répliquée sur des réseaux plus complexes, représentant idéalement la totalité des données comme dans le réseau de Turner et al, mais nos résultats confirment que les auteurs et les lecteurs doivent interpréter les méta-analyses en réseau avec prudence lorsque la question des biais de publication n'a pas été abordée.

Chapitre 6

Conclusion

Par définition, les biais de publication désignent globalement le fait de ne pas rendre publiques les informations liées à la recherche, en particulier les résultats des essais randomisés en évaluation thérapeutique. Ce n'est donc pas en premier lieu un problème statistique et les solutions doivent naturellement être recherchées en amont. Des améliorations ont bien été introduites récemment. Mis en place suite à l'initiative de l'International Committee of Medical Journal Editors, l'enregistrement des essais sur des registres publics permet ainsi de connaître l'existence des essais et savoir si les essais achevés ont été publiés (De Angelis, Drazen et al. 2004). De plus, la législation nord-américaine (FDA Amendment Act 2007) exige désormais que les résultats d'un essai (avec au moins un site aux Etats-Unis) soient postés sur clinicaltrials.gov dans l'année suivant la fin de l'essai. En Europe, suite à l'intervention du médiateur européen, l'European Medicines Agency (EMA) est maintenant tenue de fournir un accès aux rapports d'études cliniques et aux protocoles correspondants (Gotzsche and Jorgensen 2011). Cependant, le problème de la transparence n'est pas réglé et de nombreuses organisations internationales ont lancé des appels en faveur de politiques de partage de données plus strictes (Gotzsche 2011; All Trials Registered All Results Reported 2013).

La conséquence des biais de publication est que les résultats des essais rendus publics ont tendance à exagérer les effets bénéfiques des interventions de santé et de sous-estimer leurs toxicité, souvent à un degré considérable. La communauté médicale est donc potentiellement induite en erreur sur l'efficacité et l'innocuité des traitements. C'est à cette étape, où l'on cherche à synthétiser au mieux les résultats des essais, que se posent plusieurs problématiques statistiques : les essais disponibles représentent-ils un échantillon non représentatif de tous les essais existants ? les résultats de la synthèse sont-ils biaisés ? peut-on

ajuster l'estimation de l'effet du traitement sur l'effet des biais de publication et des biais connexes liés à la taille des essais ?

Ces questions concernent plus que jamais la méta-analyse en réseau. Cette méthode de synthèse est de plus en plus utilisée. Les médecins-chercheurs s'approprient de plus en plus fréquemment cette méthode car elle paraît répondre à leur préoccupation (savoir quel est le meilleur traitement disponible). Les agences d'évaluation des technologies de santé et les organismes de régulation sont également confrontés à ces méthodes de synthèse. Alors qu'il existait une littérature abondante concernant des développements statistiques, il nous est apparu que les aspects méthodologiques, en particulier le problème du biais de publication, étaient négligés dans les méta-analyses en réseau. Nous avons répondu à ce manque par l'évaluation empirique de l'impact des biais de publication et en généralisant à la méta-analyse en réseau des méthodes de détection du biais et d'analyse de sensibilité des résultats aux biais de publication.

Ces méthodes tirent leur force de l'ensemble des essais dans le réseau sous l'hypothèse d'interchangeabilité des biais, c'est-à-dire que les biais de publications s'appliquent de manière similaire sur les branches du réseau. Comme nous l'avons discuté précédemment, cette hypothèse est plausible a priori. De surcroît, on pourrait en tester la validité. Par exemple, on ajusterait un modèle de méta-régression dont les pentes seraient indépendantes; on pourrait estimer la probabilité a posteriori que la différence entre les estimations de 2 pentes quelconques soit positive.

Il est encore nécessaire d'appliquer et de valider ces méthodes pour des réseaux plus complets incluant à la fois des essais contre placebo, des essais face-face et des essais multi-bras. Comme nous l'avons décrit, les méthodes introduites permettraient techniquement de prendre en compte ces complexités. Par exemple, dans le modèle de sélection, nous pourrions distinguer la probabilité d'être publié pour un essai contre placebo de celle d'un essai face-

face; d'autre part, nous pourrions modéliser la propension à être publié d'un essai à trois bras ABC en fonction des écarts types de la comparaison AB et de la comparaison BC. Cependant, des hypothèses supplémentaires seraient nécessaires. Pour un essai face-face, les quelques données disponibles suggèrent que le biais de publication est moins fréquent pour les essais face-face que dans les essais contre placebo et la direction attendue du biais est incertaine. On pourrait supposer que le traitement sponsorisé ou que le traitement le plus récent est favorisé. Pour un essai multi-bras, le mécanisme de sélection est peu connu; on pourrait imaginer que, pour un essai comparant un traitement expérimental à un témoin actif et à un placebo, l'essai ne soit pas publié s'il ne montre pas la supériorité du traitement expérimental par rapport au placebo ou, à défaut, par rapport au témoin actif.

Nous poursuivons nos travaux de recherche sur cette thématique selon deux axes principaux. D'une part, nous cherchons à décrire une méthodologie plus générale de simulations de réseaux d'essais. Il existe peu ou pas d'études de simulation dans le cadre des méta-analyses en réseau (Lu and Ades 2009; Song, Clark et al. 2012); or elles sont nécessaires à la validation et à la comparaison des méthodes. Une telle méthodologie permettrait de générer des réseaux reflétant des géométries complexes mais réalistes (réseaux d'essais face-face et contre placebo, degrés variables de diversité et cooccurrence, information directe et indirecte cohérente ou incohérente) et d'induire des biais sur ces réseaux (par exemple de manière différentielle sur différentes parties du réseau). Nous pourrions ainsi poursuivre l'évaluation de l'impact des biais de publication (en particulier, tester des hypothèses) et la validation des méthodes d'ajustement. A cette fin, nous utiliserons les méthodes de simulation déjà développées pour évaluer le test généralisé de détection du biais. Mais nous utiliserons également un réseau étendu d'essais d'antidépresseurs, que nous avons construit, pour documenter les plans de simulation. D'autre part, il manque pour l'instant un outil graphique pour évaluer le biais, c'est-à-dire que l'on souhaiterait avoir à disposition le « funnel plot » de

la méta-analyse en réseau. Il n'est pas pertinent de superposer sur le même graphique toutes les quantités d'effet estimées θ_{ijk} par l'ensemble des essais du réseau puisque les effets vrais sous-jacents θ_{jk} varient d'une comparaison à l'autre. Par contre, on peut superposer les effets standardisés $\theta_{ijk}-\theta_{jk}$ sur un même graphique. En l'absence de biais, le nuage de points résultant devrait être symétrique autour de 0. Sous l'hypothèse de biais interchangeables, nous pourrions alors rechercher l'asymétrie sur ce graphique en entonnoir généralisé et appliquer les tests disponibles pour la méta-analyse conventionnelle, par exemple la modification du test d'Egger par Harbord, Egger et Sterne, avec une forme de stratification sur les comparaisons dans le réseau.

Bibliographie

- Ades, A. E. (2011). "ISPOR states its position on network meta-analysis." *Value Health* **14**(4): 414-416.
- Ahmed, I., A. J. Sutton, et al. (2012). "Assessment of publication bias, selection bias, and unavailable data in meta-analyses using individual participant data: a database survey." *BMJ* **344**: d7762.
- Alkhafaji, A. A., L. Trinquart, et al. (2012). "Impact of evergreening on patients and health insurance: a meta analysis and reimbursement cost analysis of citalopram/escitalopram antidepressants." *BMC Med* **10**: 142.
- All Trials Registered All Results Reported. (2013). from www.alltrials.net.
- Bafeta, A., L. Trinquart, et al. "Key methodological components of systematic reviews in network meta-analyses." *Manuscript submitted for publication*.
- Baker, S. G. and B. S. Kramer (2002). "The transitive fallacy for randomized trials: if A bests B and B bests C in separate trials, is A better than C?" *BMC Med Res Methodol* **2**: 13.
- Baujat, B., C. Mahe, et al. (2002). "A graphical method for exploring heterogeneity in meta-analyses: application to a meta-analysis of 65 trials." *Stat Med* **21**(18): 2641-2652.
- Bedrick, E. J., R. Christensen, et al. (1996). "A New Perspective on Priors for Generalized Linear Models." *JASA* **91**(436): 1450-1460.
- Begg, C. B. and M. Mazumdar (1994). "Operating characteristics of a rank correlation test for publication bias." *Biometrics* **50**(4): 1088-1101.
- Bero, L., F. Oostvogel, et al. (2007). "Factors associated with findings of published trials of drug-drug comparisons: why some statins appear more efficacious than others." *PLoS Med* **4**(6): e184.
- Borenstein, M., L. V. Hedges, et al. (2009). *Introduction to Meta-Analysis*. Chichester, UK, John Wiley & Sons, Ltd.
- Bowden, J., D. Jackson, et al. (2010). "Modelling multiple sources of dissemination bias in meta-analysis." *Stat Med* **29**(7-8): 945-955.
- Bucher, H. C., G. H. Guyatt, et al. (1997). "The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials." *J Clin Epidemiol* **50**(6): 683-691.
- Carpenter, J., G. Rucker, et al. (2011). "Assessing the sensitivity of meta-analysis to selection bias: a multiple imputation approach." *Biometrics* **67**(3): 1066-1072.
- Carpenter, J. R., G. Schwarzer, et al. (2009). "Empirical evaluation showed that the Copas selection model provided a useful summary in 80% of meta-analyses." *J Clin Epidemiol* **62**(6): 624-631 e624.
- Chaimani, A. and G. Salanti (2012). "Using network meta-analysis to evaluate the existence of small-study effects in a network of interventions." *Res Synth Methods* **3**(2): 161-176.
- Chalmers, I. (2010) "Why the 1948 MRC trial of streptomycin used treatment allocation based on random numbers." *James Lind Library*.
- Chalmers, I. and R. Matthews (2006). "What are the implications of optimism bias in clinical research?" *Lancet* **367**(9509): 449-450.
- Chan, A. W., A. Hrobjartsson, et al. (2004). "Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles." *JAMA* **291**(20): 2457-2465.
- Chatzinasiou, F., C. M. Lill, et al. (2011). "Comprehensive field synopsis and systematic meta-analyses of genetic association studies in cutaneous melanoma." *J Natl Cancer Inst* **103**(16): 1227-1235.
- Chootrakool, H., J. Q. Shi, et al. (2011). "Meta-analysis and sensitivity analysis for multi-arm trials with selection bias." *Stat Med* **30**(11): 1183-1198.
- Cipriani, A., T. A. Furukawa, et al. (2009). "Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis." *Lancet* **373**(9665): 746-758.
- Cochrane, A. L. (1972). *Effectiveness And Efficiency: Random Reflections on Health Services*, Nuffield Provincial Hospitals Trust.
- Cohen, J. (1988). *Statistical Power Analysis in the Behavioral Sciences*. Hillsdale (NJ), Lawrence Erlbaum Associates, Inc.
- Copas, J. and J. Q. Shi (2000). "Meta-analysis, funnel plots and sensitivity analysis." *Biostatistics* **1**(3): 247-262.
- Copas, J. B. and J. Q. Shi (2001). "A sensitivity analysis for publication bias in systematic reviews." *Stat Methods Med Res* **10**(4): 251-265.
- Davey, J., R. M. Turner, et al. (2011). "Characteristics of meta-analyses and their component studies in the Cochrane Database of Systematic Reviews: a cross-sectional, descriptive analysis." *BMC Med Res Methodol* **11**: 160.
- De Angelis, C., J. M. Drazen, et al. (2004). "Clinical trial registration: a statement from the International Committee of Medical Journal Editors." *Ann Intern Med* **141**(6): 477-478.
- De Angelis, C., J. M. Drazen, et al. (2004). "Clinical trial registration: a statement from the International Committee of Medical Journal Editors." *Lancet* **364**(9438): 911-912.

- DeAngelis, C. D. and P. B. Fontanarosa (2010). "Ensuring integrity in industry-sponsored research: primum non nocere, revisited." *JAMA* **303**(12): 1196-1198.
- Dear, K. B. and C. B. Begg (1992). "An Approach for Assessing Publication Bias Prior to Performing a Meta-Analysis." *Statist Sci* **7**(2): 237-245.
- Dechartres, A., I. Boutron, et al. (2011). "Single-center trials show larger treatment effects than multicenter trials: evidence from a meta-epidemiologic study." *Ann Intern Med* **155**(1): 39-51.
- DerSimonian, R. and N. Laird (1986). "Meta-analysis in clinical trials." *Control Clin Trials* **7**(3): 177-188.
- Dias, S., N. J. Welton, et al. (2010). "Study designs to detect sponsorship and other biases in systematic reviews." *J.Clin.Epidemiol.* **63**(6): 587-588.
- Dias, S., N. J. Welton, et al. (2010). "Checking consistency in mixed treatment comparison meta-analysis." *Stat Med* **29**(7-8): 932-944.
- Dias, S., N. J. Welton, et al. (2010). "Estimation and adjustment of bias in randomized evidence by using mixed treatment comparison meta-analysis." *J R Stat Soc Ser A* **173**(3): 613-629.
- Dias, S., N. J. Welton, et al. (2011). NICE DSU Technical Support Document 1: Introduction to evidence synthesis for decision making. London, National Institute for Health and Clinical Excellence.
- Dias, S., N. J. Welton, et al. (2012). NICE DSU Technical Support Document 4: Inconsistency in Networks of Evidence Based on Randomised Controlled Trials. London, National Institute for Health and Clinical Excellence.
- Dickersin, K. (2010). "Health-care policy. To reform U.S. health care, start with systematic reviews." *Science* **329**(5991): 516-517.
- Dickersin, K. and I. Chalmers (2011). "Recognising, investigating and dealing with incomplete and biased reporting of clinical research: from Francis Bacon to the World Health Organisation." *James Lind Library*.
- Dickersin, K. and D. Rennie (2003). "Registering clinical trials." *JAMA* **290**(4): 516-523.
- Donegan, S., P. Williamson, et al. (2010). "Indirect comparisons: a review of reporting and methodological quality." *PLoS One* **5**(11): e11054.
- Duval, S. and R. Tweedie (2000). "Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis." *Biometrics* **56**(2): 455-463.
- Dwan, K., D. G. Altman, et al. (2008). "Systematic review of the empirical evidence of study publication bias and outcome reporting bias." *PLoS One* **3**(8): e3081.
- Egger, M., G. Davey Smith, et al. (2001). *Systematic reviews in health care: meta-analysis in context*. London, BMJ Books.
- Egger, M., G. Davey Smith, et al. (1997). "Bias in meta-analysis detected by a simple, graphical test." *BMJ* **315**(7109): 629-634.
- Estellat, C. and P. Ravaud (2012). "Lack of head-to-head trials and fair control arms: randomized controlled trials of biologic treatment for rheumatoid arthritis." *Arch Intern Med* **172**(3): 237-244.
- Eyding, D., M. Lelgemann, et al. (2010). "Reboxetine for acute treatment of major depression: systematic review and meta-analysis of published and unpublished placebo and selective serotonin reuptake inhibitor controlled trials." *BMJ* **341**: c4737.
- Fanelli, D. (2009). "How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data." *PLoS One* **4**(5): e5738.
- Fanelli, D. (2012). "Negative results are disappearing from most disciplines and countries." *Scientometrics* **90**: 891-904.
- Fleiss, J. L., B. Levin, et al. (2004). *Statistical Methods for Rates and Proportions*. Hoboken, NJ, USA, John Wiley & Sons, Inc.
- Galbraith, R. F. (1988). "A note on graphical presentation of estimated odds ratios from several clinical trials." *Stat Med* **7**(8): 889-894.
- Gartlehner, G., B. N. Gaynes, et al. (2008). "Comparative benefits and harms of second-generation antidepressants: background paper for the American College of Physicians." *Ann Intern Med* **149**(10): 734-750.
- Gelman, A., J. B. Carlin, et al. (2004). *Bayesian Data Analysis*. New York, Chapman Hall/CRC.
- Georgiev, H., A. Tsalatsanis, et al. (2011). *Social Network Analysis of Research Programs In Multiple Myeloma*. 53rd ASH Annual Meeting and Exposition, San Diego.
- Givens, G. H., D. D. Smith, et al. (1997). "Publication bias in meta-analysis: a Bayesian data-augmentation approach to account for issues exemplified in the passive smoking debate." *Statist Sci* **12**: 221-250.
- Glenny, A. M., D. G. Altman, et al. (2005). "Indirect comparisons of competing interventions." *Health Technol.Assess.* **9**(26): 1-iv.
- Godlee, F. and T. Groves (2012). "The new BMJ policy on sharing data from drug and device trials." *BMJ* **345**: e7888.
- Godlee, F. and E. Loder (2010). "Missing clinical trial data: setting the record straight." *BMJ* **341**: c5641.

- Goldberg, N. H., S. Schneeweiss, et al. (2011). "Availability of comparative efficacy data at the time of drug approval in the United States." *JAMA* **305**(17): 1786-1789.
- Goodman, S. and K. Dickersin (2011). "Metabias: a challenge for comparative effectiveness research." *Ann Intern Med* **155**(1): 61-62.
- Gotzsche, P. C. (2005). "Research integrity and pharmaceutical industry sponsorship." *Med J Aust* **182**(11): 549-550.
- Gotzsche, P. C. (2011). "We need access to all data from all clinical trials." *Cochrane Database Syst Rev*(12): ED000035.
- Gotzsche, P. C. (2012). "Big pharma often commits corporate crime, and this must be stopped." *BMJ* **345**: e8462.
- Gotzsche, P. C., A. Hrobjartsson, et al. (2007). "Ghost authorship in industry-initiated randomised trials." *PLoS Med* **4**(1): e19.
- Gotzsche, P. C. and A. W. Jorgensen (2011). "Opening up data at the European Medicines Agency." *BMJ* **342**: d2686.
- Harbord, R. M., M. Egger, et al. (2006). "A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints." *Stat Med* **25**(20): 3443-3457.
- Hart, B., A. Lundh, et al. (2012). "Effect of reporting bias on meta-analyses of drug trials: reanalysis of meta-analyses." *BMJ* **344**: d7202.
- Hedges, L. V. (1988). "[Selection Models and the File Drawer Problem]: Comment." *Stat Sci* **3**(1): 118-120.
- Hedges, L. V. (1992). "Modeling publication selection effects in meta-analysis." *Stat Sci* **7**(2): 246-255.
- Hedges, L. V. and I. Olkin (1985). *Statistical methods for meta-analysis*. London, Academic Press.
- Hedges, L. V. and J. Vevea (2005). Selection method approaches. *Publication bias in meta-analysis*. H. Rothstein, A. J. Sutton and M. Borenstein. Chichester, John Wiley & Sons Ltd.
- Higgins, J. P., D. G. Altman, et al. (2011). "The Cochrane Collaboration's tool for assessing risk of bias in randomised trials." *BMJ* **343**: d5928.
- Higgins, J. P., J. J. Deeks, et al. (2011). Chapter 16: Special topics in statistics. *Cochrane Handbook for Systematic Reviews of Interventions*. J. P. Higgins and S. Green, The Cochrane Collaboration.
- Higgins, J. P., D. Jackson, et al. (2012). "Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies." *Res Synth Methods* **3**: 98-110.
- Higgins, J. P. and S. G. Thompson (2002). "Quantifying heterogeneity in a meta-analysis." *Stat Med* **21**(11): 1539-1558.
- Higgins, J. P., S. G. Thompson, et al. (2003). "Measuring inconsistency in meta-analyses." *BMJ* **327**(7414): 557-560.
- Higgins, J. P., S. G. Thompson, et al. (2009). "A re-evaluation of random-effects meta-analysis." *J R Stat Soc Ser A Stat Soc* **172**(1): 137-159.
- Higgins, J. P. and A. Whitehead (1996). "Borrowing strength from external trials in a meta-analysis." *Stat Med* **15**(24): 2733-2749.
- Hoaglin, D. C., N. Hawkins, et al. (2011). "Conducting indirect-treatment-comparison and network-meta-analysis studies: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 2." *Value Health* **14**(4): 429-437.
- Hochman, M. and D. McCormick (2010). "Characteristics of published comparative effectiveness studies of medications." *JAMA* **303**(10): 951-958.
- Hopewell, S., K. Loudon, et al. (2009). "Publication bias in clinical trials due to statistical significance or direction of trial results." *Cochrane Database Syst.Rev.*(1): MR000006.
- Hrynaskiewicz, I. and D. G. Altman (2009). "Towards agreement on best practice for publishing raw clinical trial data." *Trials* **10**: 17.
- Ioannidis, J. P. (2005). "Why most published research findings are false." *PLoS Med* **2**(8): e124.
- Ioannidis, J. P. (2008). "Effectiveness of antidepressants: an evidence myth constructed from a thousand randomized trials?" *Philos Ethics Humanit Med* **3**: 14.
- Ioannidis, J. P. (2008). "Interpretation of tests of heterogeneity and bias in meta-analysis." *J Eval Clin Pract* **14**(5): 951-957.
- Ioannidis, J. P. (2009). "Integration of evidence from multiple meta-analyses: a primer on umbrella reviews, treatment networks and multiple treatments meta-analyses." *CMAJ* **181**(8): 488-493.
- Ioannidis, J. P. (2011). "Excess significance bias in the literature on brain volume abnormalities." *Arch Gen Psychiatry* **68**(8): 773-780.
- Ioannidis, J. P. and T. A. Trikalinos (2007). "The appropriateness of asymmetry tests for publication bias in meta-analyses: a large survey." *CMAJ* **176**(8): 1091-1096.
- Ioannidis, J. P. and T. A. Trikalinos (2007). "Authors' response to V Johnson and Y Yuan." *Clin Trials* **4**(3): 256-257.

- Ioannidis, J. P. and T. A. Trikalinos (2007). "An exploratory test for an excess of significant findings." *Clin Trials* **4**(3): 245-253.
- Iom (2009). *Initial National Priorities for Comparative Effectiveness Research*. Washington, DC, The National Academies Press.
- Iyengar, S. and J. B. Greenhouse (1988). "Selection Models and the File Drawer Problem." *Statist Sci* **3**(1): 109-117.
- Iyengar, S. and J. B. Greenhouse (1988). "Selection models and the file drawer problem (with discussion)." *Statist Sci* **3**: 109- 135.
- Jackson, D. (2007). "Assessing the implications of publication bias for two popular estimates of between-study variance in meta-analysis." *Biometrics* **63**(1): 187-193.
- Jansen, J. P., R. Fleurence, et al. (2011). "Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 1." *Value Health* **14**(4): 417-428.
- Jones, B., J. Roger, et al. (2011). "Statistical approaches for conducting network meta-analysis in drug development." *Pharm Stat* **10**(6): 523-531.
- Kavvoura, F. K., M. B. McQueen, et al. (2008). "Evaluation of the potential excess of statistically significant findings in published genetic association studies: application to Alzheimer's disease." *Am J Epidemiol* **168**(8): 855-865.
- Kent, D. M. and D. E. Thaler (2008). "Stroke prevention--insights from incoherence." *N Engl J Med* **359**(12): 1287-1289.
- L'Abbe, K. A., A. S. Detsky, et al. (1987). "Meta-analysis in clinical research." *Ann Intern Med* **107**(2): 224-233.
- Lambert, P. C., A. J. Sutton, et al. (2005). "How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS." *Stat Med* **24**(15): 2401-2428.
- Langan, D., J. P. Higgins, et al. (2012). "Graphical augmentations to the funnel plot assess the impact of additional evidence on a meta-analysis." *J Clin Epidemiol* **65**(5): 511-519.
- Larose, D. T. and D. K. Dey (1998). "Modeling publication bias using weighted distributions in a Bayesian framework." *Comput Stat Data Anal* **26**: 279-302.
- Lathyris, D. N., N. A. Patsopoulos, et al. (2010). "Industry sponsorship and selection of comparators in randomized clinical trials." *Eur J Clin Invest* **40**(2): 172-182.
- Lee, K., P. Bacchetti, et al. (2008). "Publication of clinical trials supporting successful new drug applications: a literature analysis." *PLoS Med* **5**(9): e191.
- Lexchin, J. (1993). "Interactions between physicians and the pharmaceutical industry: what does the literature say?" *CMAJ* **149**(10): 1401-1407.
- Lexchin, J. (2012). "Those who have the gold make the evidence: how the pharmaceutical industry biases the outcomes of clinical trials of medications." *Sci Eng Ethics* **18**(2): 247-261.
- Lexchin, J., L. A. Bero, et al. (2003). "Pharmaceutical industry sponsorship and research outcome and quality: systematic review." *BMJ* **326**(7400): 1167-1170.
- Li, T., M. A. Puhan, et al. (2011). "Network meta-analysis-highly attractive but more methodological research is needed." *BMC Med* **9**: 79.
- Lloyd, C. J. (2005). "Estimating test power adjusted for size." *J Stat Comput Simulat* **75**(11): 921-933.
- Lu, G. and A. Ades (2009). "Modeling between-trial variance structure in mixed treatment comparisons." *Biostatistics* **10**(4): 792-805.
- Lu, G. and A. E. Ades (2004). "Combination of direct and indirect evidence in mixed treatment comparisons." *Stat Med* **23**(20): 3105-3124.
- Lu, G. and A. E. Ades (2006). "Assessing Evidence Inconsistency in Mixed Treatment Comparisons." *JASA* **101**(474): 447-459.
- Lumley, T. (2002). "Network meta-analysis for indirect treatment comparisons." *Stat Med* **21**(16): 2313-2324.
- Lundh, A., S. Sismondo, et al. (2012). "Industry sponsorship and research outcome." *Cochrane Database Syst Rev* **12**: MR000033.
- Macaskill, P., S. D. Walter, et al. (2001). "A comparison of methods to detect publication bias in meta-analysis." *Stat Med* **20**(4): 641-654.
- Mathieu, S., I. Boutron, et al. (2009). "Comparison of registered and published primary outcomes in randomized controlled trials." *JAMA* **302**(9): 977-984.
- Mauri, D., N. P. Polyzos, et al. (2008). "Multiple-treatments meta-analysis of chemotherapy and targeted therapies in advanced breast cancer." *J Natl Cancer Inst* **100**(24): 1780-1791.
- Mavridis, D., A. Sutton, et al. (2012). "A fully Bayesian application of the Copas selection model for publication bias extended to network meta-analysis." *Stat Med*.
- McCandless, L. C., P. Gustafson, et al. (2012). "Hierarchical priors for bias parameters in Bayesian sensitivity analysis for unmeasured confounding." *Stat Med* **31**(4): 383-396.

- McGauran, N., B. Wieseler, et al. (2010). "Reporting bias in medical research - a narrative review." *Trials* **11**: 37.
- Miller, J. D. (2010). "Registering Clinical Trial Results." *JAMA* **303**(8): 773-774.
- Mills, E. J., J. P. Ioannidis, et al. (2012). "How to use an article reporting a multiple treatment comparison meta-analysis." *JAMA* **308**(12): 1246-1253.
- Montori, V., J. P. Ioannidis, et al. (2008). Reporting bias. *User's guide to the medical literature: a manual for evidence-based clinical practice*. G. H. Guyatt, D. Rennie, M. O. Meade and D. J. Cook, The McGraw-Hill Companies. **Second**: 543-554.
- Moreno, S. G., A. J. Sutton, et al. (2011). "Adjusting for publication biases across similar interventions performed well when compared with gold standard data." *J Clin Epidemiol* **64**(11): 1230-1241.
- Moreno, S. G., A. J. Sutton, et al. (2009). "Assessment of regression-based methods to adjust for publication bias through a comprehensive simulation study." *BMC.Med Res.Methodol* **9**: 2.
- Moreno, S. G., A. J. Sutton, et al. (2012). "A generalized weighting regression-derived meta-analysis estimator robust to small-study effects and heterogeneity." *Stat Med*.
- Moreno, S. G., A. J. Sutton, et al. (2009). "Novel methods to deal with publication biases: secondary analysis of antidepressant trials in the FDA trial registry database and related journal publications." *BMJ* **339**: b2981.
- Ntzani, E. E., E. C. Rizos, et al. (2007). "Genetic effects versus bias for candidate polymorphisms in myocardial infarction: case study and overview of large-scale evidence." *Am J Epidemiol* **165**(9): 973-984.
- Nuesch, E., S. Trelle, et al. (2010). "Small study effects in meta-analyses of osteoarthritis trials: meta-epidemiological study." *BMJ* **341**: c3515.
- O'Connor, A. B. (2009). "The Need for Improved Access to FDA Reviews." *JAMA* **302**(2): 191-193.
- O'Regan, C., I. Ghement, et al. (2009). "Incorporating multiple interventions in meta-analysis: an evaluation of the mixed treatment comparison with the adjusted indirect comparison." *Trials* **10**: 86.
- Pan, Z., T. A. Trikalinos, et al. (2005). "Local literature bias in genetic epidemiology: an empirical evaluation of the Chinese literature." *PLoS Med* **2**(12): e334.
- Peters, J. L., A. J. Sutton, et al. (2006). "Comparison of two methods to detect publication bias in meta-analysis." *JAMA* **295**(6): 676-680.
- Peters, J. L., A. J. Sutton, et al. (2008). "Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry." *J Clin Epidemiol* **61**(10): 991-996.
- Preston, C., D. Ashby, et al. (2004). "Adjusting for publication bias: modelling the selection process." *J Eval Clin Pract* **10**(2): 313-322.
- Riley, R. D., J. P. Higgins, et al. (2011). "Interpretation of random effects meta-analyses." *BMJ* **342**: d549.
- Riley, R. D., P. C. Lambert, et al. (2010). "Meta-analysis of individual participant data: rationale, conduct, and reporting." *BMJ* **340**: c221.
- Rising, K., P. Bacchetti, et al. (2008). "Reporting bias in drug trials submitted to the Food and Drug Administration: review of publication and presentation." *PLoS.Med* **5**(11): e217.
- Rizos, E. C., G. Salanti, et al. (2011). "Homophily and co-occurrence patterns shape randomized trials agendas: illustration in antifungal agents." *J Clin Epidemiol* **64**(8): 830-842.
- Rothstein, H. R., A. J. Sutton, et al. (2005). *Publication bias in meta-analysis : prevention, assessment and adjustments*. Chichester, UK, John Wiley & Sons Ltd.
- Rucker, G., J. R. Carpenter, et al. (2011). "Detecting and adjusting for small-study effects in meta-analysis." *Biom J* **53**(2): 351-368.
- Rucker, G., G. Schwarzer, et al. (2008). "Arcsine test for publication bias in meta-analyses with binary outcomes." *Stat Med* **27**(5): 746-763.
- Rucker, G., G. Schwarzer, et al. (2011). "Treatment-effect estimates adjusted for small-study effects via a limit meta-analysis." *Biostatistics* **12**(1): 122-142.
- Rufibach, K. (2011). "Selection models with monotone weight functions in meta analysis." *Biom J* **53**(4): 689-704.
- Salanti, G., A. E. Ades, et al. (2010). "Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial." *J.Clin.Epidemiol.* **64**(2): 163-171.
- Salanti, G., S. Dias, et al. (2010). "Evaluating novel agent effects in multiple-treatments meta-regression." *Stat Med* **29**(23): 2369-2383.
- Salanti, G., J. P. T. Higgins, et al. (2008). "Evaluation of networks of randomized trials." *Stat Methods Med Res* **17**(3): 279-301.
- Salanti, G., F. K. Kavvoura, et al. (2008). "Exploring the geometry of treatment networks." *Ann Intern Med* **148**(7): 544-553.
- Salanti, G., V. Marinho, et al. (2009). "A case study of multiple-treatments meta-analysis demonstrates that covariates should be considered." *J.Clin.Epidemiol.* **62**(8): 857-864.

- Savovic, J., H. E. Jones, et al. (2012). "Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials." *Ann Intern Med* **157**(6): 429-438.
- Schöttker, B., D. Luhmann, et al. (2009). Indirekte Vergleiche von Therapieverfahren. Köln, Deutschen Institut für Medizinische Dokumentation und Information.
- Schwarzer, G., G. Antes, et al. (2002). "Inflation of type I error rate in two statistical tests for the detection of publication bias in meta-analyses with binary outcomes." *Stat Med* **21**(17): 2465-2477.
- Schwarzer, G., G. Antes, et al. (2007). "A test for publication bias in meta-analysis with sparse binary data." *Stat Med* **26**(4): 721-733.
- Schwarzer, G., J. Carpenter, et al. (2010). "Empirical evaluation suggests Copas selection model preferable to trim-and-fill method for selection bias in meta-analysis." *J Clin Epidemiol* **63**(3): 282-288.
- Senn, S., F. Gavini, et al. (2012). "Issues in performing a network meta-analysis." *Stat Methods Med Res*.
- Sidik, K. and J. N. Jonkman (2005). "Simple heterogeneity variance estimation for meta-analysis." *J R Stat Soc Ser C* **54**: 367-384.
- Sidik, K. and J. N. Jonkman (2007). "A comparison of heterogeneity variance estimators in combining results of studies." *Stat Med* **26**: 1964-1981.
- Silliman, N. P. (1997). "Hierarchical Selection Models With Applications in Meta-Analysis." *JASA* **92**(439): 926-936.
- Smith, T. C., D. J. Spiegelhalter, et al. (1995). "Bayesian approaches to random-effects meta-analysis: a comparative study." *Stat Med* **14**(24): 2685-2699.
- Smyth, R. M., J. J. Kirkham, et al. (2011). "Frequency and reasons for outcome reporting bias in clinical trials: interviews with trialists." *BMJ* **342**: c7153.
- Song, F., D. G. Altman, et al. (2003). "Validity of indirect comparison for estimating efficacy of competing interventions: empirical evidence from published meta-analyses." *BMJ* **326**(7387): 472.
- Song, F., A. Clark, et al. (2012). "Simulation evaluation of statistical properties of methods for indirect and mixed treatment comparisons." *BMC Med Res Methodol* **12**: 138.
- Song, F., A. J. Eastwood, et al. (2000). "Publication and related biases." *Health Technol Assess* **4**(10): 1-115.
- Song, F., I. Harvey, et al. (2008). "Adjusted indirect comparison may be less biased than direct comparison for evaluating new pharmaceutical interventions." *J Clin Epidemiol* **61**(5): 455-463.
- Song, F., Y. K. Loke, et al. (2009). "Methodological problems in the use of indirect comparisons for evaluating healthcare interventions: survey of published systematic reviews." *BMJ* **338**: b1147.
- Song, F., S. Parekh-Bhurke, et al. (2009). "Extent of publication bias in different categories of research cohorts: a meta-analysis of empirical studies." *BMC Med Res Methodol* **9**: 79.
- Song, F., S. Parekh, et al. (2010). "Dissemination and publication of research findings: an updated review of related biases." *Health Technol Assess* **14**(8): iii, ix-iii, 193.
- Song, F., S. Parekh, et al. (2010). "Dissemination and publication of research findings: an updated review of related biases." *Health Technol Assess* **14**(8): iii, ix-xi, 1-193.
- Song, F., T. Xiong, et al. (2011). "Inconsistency between direct and indirect comparisons of competing interventions: meta-epidemiological study." *BMJ* **343**: d4909.
- Spiegelhalter, D. J. (1998). "Bayesian graphical modelling: a case study in monitoring health outcomes." *J R Stat Soc Ser C* **47**: 115-133.
- Spiegelhalter, D. J., N. G. Best, et al. (2002). "Bayesian measures of model complexity and fit." *J R Stat Soc Ser B* **64**(4): 583-639.
- Stafford, R. S., T. H. Wagner, et al. (2009). "New, but not improved? Incorporating comparative-effectiveness information into FDA labeling." *N Engl J Med* **361**(13): 1230-1233.
- Stanley, T. D. (2008). "Meta-Regression Methods for Detecting and Estimating Empirical Effects in the Presence of Publication Selection*." *Oxf Bull Econ Stat* **70**(1): 103-127.
- Sterne, J., M. Egger, et al. (2011). Chapter 10: Addressing reporting biases. *Cochrane Handbook for Systematic Reviews of Interventions*. J. P. Higgins and S. Green, The Cochrane Collaboration.
- Sterne, J. A., M. Egger, et al. (2001). "Systematic reviews in health care: Investigating and dealing with publication and other biases in meta-analysis." *BMJ* **323**(7304): 101-105.
- Sterne, J. A., A. J. Sutton, et al. (2011). "Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials." *BMJ* **343**: d4002.
- Stijnen, T., T. H. Hamza, et al. (2010). "Random effects meta-analysis of event outcome in the framework of the generalized linear mixed model with applications in sparse data." *Stat Med* **29**(29): 3046-3067.
- Sutton, A. J., S. J. Duval, et al. (2000). "Empirical assessment of effect of publication bias on meta-analyses." *BMJ* **320**(7249): 1574-1577.
- Sutton, A. J., F. Song, et al. (2000). "Modelling publication bias in meta-analysis: a review." *Stat Methods Med Res* **9**(5): 421-445.
- Terrin, N., C. H. Schmid, et al. (2005). "In an empirical evaluation of the funnel plot, researchers could not visually identify publication bias." *J Clin Epidemiol* **58**(9): 894-901.

- Thompson, S. G. and J. P. Higgins (2002). "How should meta-regression analyses be undertaken and interpreted?" *Stat Med* **21**(11): 1559-1573.
- Thompson, S. G. and S. J. Sharp (1999). "Explaining heterogeneity in meta-analysis: a comparison of methods." *Stat Med* **18**(20): 2693-2708.
- Thorlund, K., L. Thabane, et al. (2013). "Modelling heterogeneity variances in multiple treatment comparison meta-analysis--are informative priors the better solution?" *BMC Med Res Methodol* **13**: 2.
- Trinquart, L., A. Abbé, et al. (2012). "Impact of Reporting Bias in Network Meta-Analysis of Antidepressant Placebo-Controlled Trials." *PLoS One* **7**(4): e35219.
- Trinquart, L., G. Chatellier, et al. (2012). "Adjustment for reporting bias in network meta-analysis of antidepressant trials." *BMC Med Res Methodol* **12**(1): 150.
- Tsilidis, K. K., S. I. Papatheodorou, et al. (2012). "Evaluation of Excess Statistical Significance in Meta-analyses of 98 Biomarker Associations with Cancer Risk." *J Natl Cancer Inst* **104**(24): 1867-1878.
- Turner, E. (2005). "Correction/Clarification about FDA Review Documents." *PLoS Med* **2**(12): e422.
- Turner, E. H. (2004). "A taxpayer-funded clinical trials registry and results database." *PLoS Med* **1**(3): e60.
- Turner, E. H., D. Knoepfelmacher, et al. (2012). "Publication bias in antipsychotic trials: an analysis of efficacy comparing the published literature to the US Food and Drug Administration database." *PLoS Med* **9**(3): e1001189.
- Turner, E. H., A. M. Matthews, et al. (2008). "Selective publication of antidepressant trials and its influence on apparent efficacy." *N Engl J Med* **358**(3): 252-260.
- Turner, R. M., J. Davey, et al. (2012). "Predicting the extent of heterogeneity in meta-analysis, using empirical data from the Cochrane Database of Systematic Reviews." *Int J Epidemiol* **41**(3): 818-827.
- Tzoulaki, I., K. C. Siontis, et al. ((in press)). "Bias in associations of emerging biomarkers with cardiovascular disease." *JAMA Internal Medicine*.
- U.S. National Institutes of Health ClinicalTrials.gov Protocol registration system.
- Vedula, S. S., L. Bero, et al. (2009). "Outcome reporting in industry-sponsored trials of gabapentin for off-label use." *N Engl J Med* **361**(20): 1963-1971.
- Vevea, J. L. and L. V. Hedges (1995). "A general linear model for estimating effect size in the presence of publication bias." *Psychometrika* **60**: 419-435.
- Vevea, J. L. and C. M. Woods (2005). "Publication bias in research synthesis: sensitivity analysis using a priori weight functions." *Psychol Methods* **10**: 428-443.
- Volpp, K. G. and A. Das (2009). "Comparative effectiveness--thinking beyond medication A versus medication B." *N Engl J Med* **361**(4): 331-333.
- Wells, G. A., S. A. Sultan, et al. (2009). Indirect Evidence: Indirect Treatment Comparisons in Meta-Analysis. Ottawa, Canadian Agency for Drugs and Technologies in Health.
- Welton, N. J., D. M. Caldwell, et al. (2009). "Mixed treatment comparison meta-analysis of complex interventions: psychological interventions in coronary heart disease." *Am.J.Epidemiol.* **169**(9): 1158-1165.
- White, I. R., J. K. Barrett, et al. (2012). "Consistency and inconsistency in network meta-analysis: model estimation using multivariate meta-regression." *Res Synth Methods* **3**: 111-125.
- Whitehead, A. (2002). *Meta-analysis of Controlled Clinical Trials*. Chichester, UK, John Wiley & Sons, Ltd.
- World Health Organization International Clinical Trials Registry Platform.
- Wu, Y., W. J. Shih, et al. (2008). "Elicitation of a beta prior for Bayesian inference in clinical trials." *Biom J* **50**(2): 212-223.
- Xiong, T., S. Parekh-Bhurke, et al. (2013). "Overall similarity and consistency assessment scores are not sufficiently accurate for predicting discrepancy between direct and indirect comparison estimates." *J Clin Epidemiol* **66**(2): 184-191.
- Zarin, D. A. and T. Tse (2008). "Medicine. Moving toward transparency of clinical trials." *Science* **319**(5868): 1340-1342.
- Zarin, D. A., T. Tse, et al. (2011). "The ClinicalTrials.gov results database--update and key issues." *N Engl J Med* **364**(9): 852-860.

Annexes

Annexe 1, Texte 1: Description du modèle

We used a mixed treatment comparison meta-analysis model, as described by Lu and Ades [7], which is a generalization of a standard pair-wise meta-analysis. The model originally concerned binary data, but because it was modeled on a logit scale as continuous data, it can be adapted to continuous outcomes. Indeed, application of this model for study-level continuous data, based on the standardized mean difference (SMD) as the scale of measurement, was outlined previously [22;23].

For each randomized controlled trial (RCT) j , the data provide information on φ_{jk} , the difference between antidepressant agent k and placebo in mean change from baseline in depression score. Because of the 2 different scales used to assess depression, the φ_{jk} s were standardized by the pooled standard deviation within groups, $\delta_{jk} = \varphi_{jk} / SD_{jk}$; that is, δ_{jk} is the SMD between agent k and placebo. Moreover, Hedges' correction for small sample size was applied in that the effect measure was multiplied by a correction factor.

Note that the placebo is the control for every RCT and is considered the reference for the whole network.

The model is a random-effects hierarchical Bayesian model, which can be written as follows. δ_{jk} is the observed effect size of antidepressant agent k versus placebo in RCT j . It is described by the study-specific true effect size d_{jk} and sampling variance ξ_j^2 according to a normal distribution:

$$\delta_{jk} \sim \text{Normal} (d_{jk} , \xi_j^2)$$

Because we chose a random effects model, the true study-specific effect size, d_{jk} , is assumed to be normally distributed:

$$d_{jk} \sim \text{Normal} (\lambda_k , \tau^2)$$

The basic parameters of the model are the mean effect size λ_k and between-study variance τ^2 , which describes the extent of the heterogeneity among true study-specific effect sizes. We applied an assumption of homogeneous variance, $\tau_{jk}^2 = \tau^2$. If τ^2 is 0, there is no between-study heterogeneity, for a fixed-effects model.

Because we chose a Bayesian model, we defined prior distributions for λ_k and τ^2 . We chose vague prior distributions for λ_k (i.e., normal distributions with mean 0 and a large variance) and a weakly informative prior distribution for the between-study heterogeneity, namely, a flat uniform distribution:

$$\lambda_k \sim \text{Normal} (0, 1000)$$

$$\tau^2 \sim \text{Uniform} (0, 2)$$

The model was fitted by use of Bayesian inference computed with Monte Carlo Markov chain simulation. The model was implemented by use of WinBUGS v1.4.3 (Imperial College and Medical Research Council, 2004) via the R2WinBUGS package in R software (R Development Core Team, v2.11.1, Vienna, Austria). Convergence was assessed by use of the Brooks-Gelman-Rubin diagnostic. We constructed posterior distributions of the SMDs from 3 chains of 80,000 simulations, after convergence was achieved from an initial 8,000 simulations for each (burn-in). For each antidepressant agent, we estimated the SMD against the placebo (basic parameter). We derived the 66 comparisons between antidepressant agents (functional parameter) from differences between basic parameters. For each pair-wise comparison, we estimated the median SMD and its associated Bayesian 95% credible interval. Statistical significance was achieved at the 5% level when the 95% credibility interval did not include 0. In each simulation, we ranked the best treatment option with the highest SMD. Ranking of the competing drugs was assessed using the median of the posterior distribution for the rank of each drug and the probability that each treatment was best was derived from the percentage of the best ranking across all simulations.

Annexe 1, Texte 2: Code Winbugs

Observed data are delta, the study-specific SMD, with s.e. zeta

```
model{
for(j in 1:J) {
  p[j]<-1/pow(zeta[j],2)
  delta[j]~dnorm(d[j],p[j])
  d[j]~dnorm(lambda[t[j]],tau)
}
lambda[1]<-0
```

```
# vague priors for NT-1 basic parameters
for (k in 2:NT) { lambda[k] ~ dnorm(0,1.0E-3) }

# vague prior for random-effect heterogeneity
sigma~dunif(0,2)
tau<-1/ pow(sigma,2)

# pair-wise SMDs
for (k in 1:(NT-1)) {
  for (l in (k+1):NT) { smd[k,l] <- lambda[l] - lambda[k] }
}

# Ranking and Prob{treatment k is best}
n<-seq(to=NT, from=1)
for (m in 1: NT){
  rk[m]<-NT+1-rank(lambda[,m])
  for (k in 1:NT) { best[k, n[m]]<-equals(rk[k],n[m]) }
}
```

Annexe 1, Tableau 1: Quantités d'effet estimés par les méta-analyses en réseau

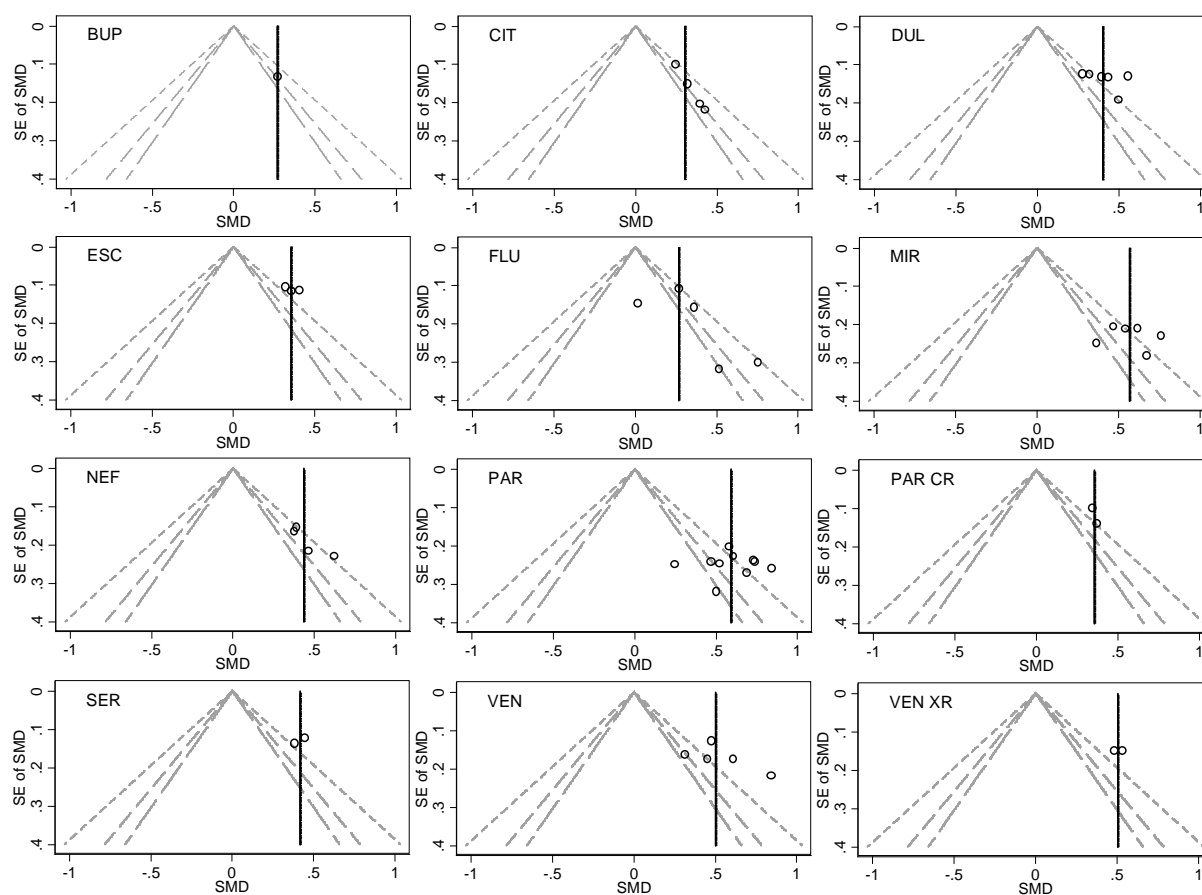
agent A versus B		74 essais avec quantités d'effet issues des rapports d'évaluation par la FDA		51 essais avec quantités d'effet issues des articles publiés		51 essais avec quantités d'effet issues des rapports d'évaluation par la FDA	
		médiane [intervalle de crédibilité]	P(A>B)	médiane [intervalle de crédibilité]	P(A>B)	médiane [intervalle de crédibilité]	P(A>B)
CIT	BUP	0.064 [-0.152; 0.278]	71.9	0.035 [-0.274; 0.341]	58.7	0.032 [-0.284; 0.347]	57.8
DUL	BUP	0.123 [-0.069; 0.315]	89.6	0.132 [-0.165; 0.422]	80.9	0.073 [-0.225; 0.371]	68.4
ESC	BUP	0.132 [-0.075; 0.34]	89.5	0.087 [-0.216; 0.388]	71.3	0.081 [-0.226; 0.389]	69.9
FLU	BUP	0.079 [-0.144; 0.307]	75.8	0.001 [-0.309; 0.308]	50.1	-0.015 [-0.327; 0.300]	46.4
MIR	BUP	0.174 [-0.036; 0.384]	94.8	0.297 [-0.033; 0.624]	96.2	0.177 [-0.143; 0.500]	86.1
NEF	BUP	0.082 [-0.136; 0.302]	76.9	0.167 [-0.162; 0.495]	84.0	0.061 [-0.263; 0.385]	64.4
PAR	BUP	0.249 [0.048; 0.452]	99.2	0.322 [0.012; 0.631]	97.9	0.279 [-0.037; 0.593]	96.0
PAR CR	BUP	0.146 [-0.107; 0.402]	87.1	0.085 [-0.234; 0.401]	69.8	0.050 [-0.282; 0.386]	61.8
SER	BUP	0.075 [-0.146; 0.296]	74.8	0.149 [-0.183; 0.476]	81.2	0.029 [-0.299; 0.360]	57.1
VEN	BUP	0.217 [0.006; 0.429]	97.9	0.235 [-0.075; 0.542]	93.2	0.175 [-0.137; 0.487]	86.5
VEN XR	BUP	0.219 [-0.024; 0.464]	96.1	0.238 [-0.11; 0.578]	91.3	0.246 [-0.104; 0.597]	91.6
DUL	CIT	0.059 [-0.119; 0.24]	74.2	0.096 [-0.091; 0.283]	84.3	0.041 [-0.149; 0.233]	66.4
ESC	CIT	0.069 [-0.126; 0.266]	75.5	0.052 [-0.147; 0.250]	69.5	0.049 [-0.155; 0.256]	68.3
FLU	CIT	0.016 [-0.198; 0.233]	56.0	-0.034 [-0.245; 0.174]	37.4	-0.046 [-0.260; 0.168]	33.7
MIR	CIT	0.111 [-0.087; 0.311]	86.2	0.263 [0.025; 0.498]	98.5	0.145 [-0.080; 0.373]	89.6
NEF	CIT	0.018 [-0.189; 0.229]	56.9	0.131 [-0.102; 0.368]	86.4	0.029 [-0.201; 0.258]	59.8
PAR	CIT	0.185 [-0.003; 0.379]	97.2	0.287 [0.073; 0.501]	99.6	0.247 [0.033; 0.462]	98.8
PAR CR	CIT	0.083 [-0.162; 0.333]	74.6	0.049 [-0.173; 0.273]	66.9	0.019 [-0.222; 0.261]	56.3
SER	CIT	0.011 [-0.197; 0.222]	54.3	0.114 [-0.127; 0.351]	82.4	-0.002 [-0.24; 0.237]	49.5
VEN	CIT	0.154 [-0.045; 0.356]	93.4	0.200 [-0.01; 0.407]	96.9	0.144 [-0.068; 0.353]	91.0

agent A versus B		74 essais avec quantités d'effet issues des rapports d'évaluation par la FDA		51 essais avec quantités d'effet issues des articles publiés		51 essais avec quantités d'effet issues des rapports d'évaluation par la FDA	
		médiane [intervalle de crédibilité]	P(A>B)	médiane [intervalle de crédibilité]	P(A>B)	médiane [intervalle de crédibilité]	P(A>B)
VEN XR	CIT	0.155 [-0.077; 0.392]	90.3	0.203 [-0.056; 0.461]	93.9	0.215 [-0.053; 0.484]	94.2
ESC	DUL	0.010 [-0.160; 0.180]	54.4	-0.044 [-0.219; 0.129]	30.8	0.009 [-0.169; 0.185]	53.8
DUL	FLU	0.043 [-0.149; 0.233]	67.0	0.131 [-0.054; 0.314]	91.7	0.086 [-0.099; 0.276]	82.0
MIR	DUL	0.052 [-0.121; 0.225]	71.9	0.166 [-0.049; 0.383]	93.3	0.104 [-0.098; 0.307]	84.5
DUL	NEF	0.041 [-0.145; 0.223]	66.8	-0.035 [-0.25; 0.179]	37.2	0.012 [-0.194; 0.219]	54.5
PAR	DUL	0.126 [-0.036; 0.291]	93.6	0.191 [0.001; 0.380]	97.6	0.206 [0.018; 0.394]	98.4
PAR CR	DUL	0.024 [-0.203; 0.250]	58.2	-0.047 [-0.247; 0.155]	32.5	-0.022 [-0.240; 0.196]	42.2
DUL	SER	0.048 [-0.137; 0.233]	69.4	-0.017 [-0.235; 0.200]	43.7	0.043 [-0.171; 0.255]	65.3
VEN	DUL	0.095 [-0.080; 0.271]	85.5	0.104 [-0.081; 0.288]	86.5	0.103 [-0.080; 0.286]	86.4
VEN XR	DUL	0.096 [-0.116; 0.309]	81.3	0.107 [-0.133; 0.346]	80.9	0.174 [-0.073; 0.421]	91.6
ESC	FLU	0.052 [-0.155; 0.259]	69.1	0.086 [-0.112; 0.285]	80.2	0.096 [-0.106; 0.298]	82.4
MIR	ESC	0.042 [-0.148; 0.232]	66.7	0.210 [-0.015; 0.439]	96.5	0.096 [-0.119; 0.311]	80.9
ESC	NEF	0.051 [-0.151; 0.249]	68.9	-0.081 [-0.306; 0.146]	24.3	0.020 [-0.197; 0.238]	57.4
PAR	ESC	0.116 [-0.063; 0.300]	89.8	0.236 [0.031; 0.438]	98.8	0.198 [-0.005; 0.400]	97.2
PAR CR	ESC	0.014 [-0.224; 0.255]	54.7	-0.002 [-0.214; 0.212]	49.2	-0.03 [-0.261; 0.200]	39.7
ESC	SER	0.057 [-0.143; 0.260]	71.3	-0.061 [-0.292; 0.167]	29.8	0.051 [-0.175; 0.276]	67.2
VEN	ESC	0.085 [-0.107; 0.279]	80.9	0.149 [-0.051; 0.347]	92.9	0.094 [-0.104; 0.292]	82.5
VEN XR	ESC	0.086 [-0.138; 0.315]	77.4	0.151 [-0.099; 0.402]	88.2	0.165 [-0.092; 0.424]	89.5
MIR	FLU	0.094 [-0.117; 0.304]	81.0	0.296 [0.061; 0.532]	99.3	0.191 [-0.033; 0.417]	95.3
NEF	FLU	0.002 [-0.216; 0.219]	50.7	0.165 [-0.068; 0.401]	91.8	0.075 [-0.152; 0.302]	74.1
PAR	FLU	0.169 [-0.032; 0.370]	95.1	0.322 [0.110; 0.532]	99.9	0.293 [0.080; 0.504]	99.7
PAR CR	FLU	0.067 [-0.188; 0.323]	69.6	0.084 [-0.138; 0.306]	77.1	0.065 [-0.174; 0.303]	70.4
FLU	SER	0.005 [-0.214; 0.225]	51.8	-0.148 [-0.384; 0.09]	11.1	-0.044 [-0.279; 0.191]	35.5
VEN	FLU	0.138 [-0.073; 0.348]	90.0	0.234 [0.026; 0.443]	98.6	0.189 [-0.017; 0.398]	96.3

agent A versus B		74 essais avec quantités d'effet issues des rapports d'évaluation par la FDA		51 essais avec quantités d'effet issues des articles publiés		51 essais avec quantités d'effet issues des rapports d'évaluation par la FDA	
		médiane [intervalle de crédibilité]	P(A>B)	médiane [intervalle de crédibilité]	P(A>B)	médiane [intervalle de crédibilité]	P(A>B)
VEN XR	FLU	0.139 [-0.103; 0.382]	87.0	0.237 [-0.021; 0.493]	96.4	0.260 [-0.004; 0.529]	97.3
MIR	NEF	0.093 [-0.112; 0.295]	81.3	0.131 [-0.129; 0.389]	83.7	0.116 [-0.122; 0.357]	83.0
PAR	MIR	0.075 [-0.109; 0.260]	78.8	0.026 [-0.212; 0.262]	58.2	0.101 [-0.122; 0.326]	81.2
MIR	PAR CR	0.028 [-0.214; 0.270]	58.8	0.212 [-0.035; 0.461]	95.4	0.126 [-0.122; 0.377]	84.0
MIR	SER	0.099 [-0.105; 0.304]	82.9	0.148 [-0.113; 0.413]	86.6	0.148 [-0.099; 0.393]	88.0
VEN	MIR	0.044 [-0.152; 0.240]	66.9	-0.061 [-0.299; 0.174]	30.4	-0.002 [-0.222; 0.218]	49.3
VEN XR	MIR	0.045 [-0.183; 0.273]	64.9	-0.058 [-0.34; 0.218]	33.9	0.069 [-0.208; 0.348]	68.7
PAR	NEF	0.167 [-0.030; 0.362]	95.3	0.156 [-0.084; 0.394]	90.0	0.218 [-0.009; 0.444]	97.0
PAR CR	NEF	0.064 [-0.183; 0.314]	69.5	-0.081 [-0.331; 0.163]	25.8	-0.01 [-0.263; 0.242]	46.9
NEF	SER	0.007 [-0.207; 0.223]	52.6	0.018 [-0.242; 0.277]	55.5	0.031 [-0.219; 0.279]	59.5
VEN	NEF	0.136 [-0.069; 0.341]	90.3	0.068 [-0.167; 0.303]	71.5	0.115 [-0.112; 0.337]	84.2
VEN XR	NEF	0.137 [-0.101; 0.374]	87.2	0.071 [-0.209; 0.349]	69.1	0.186 [-0.095; 0.467]	90.4
PAR	PAR CR	0.103 [-0.132; 0.337]	80.5	0.237 [0.010; 0.464]	98.0	0.228 [-0.011; 0.466]	97.0
PAR	SER	0.174 [-0.019; 0.371]	96.0	0.174 [-0.066; 0.413]	92.2	0.249 [0.014; 0.485]	98.1
PAR	VEN	0.031 [-0.154; 0.218]	63.1	0.087 [-0.124; 0.297]	79.1	0.104 [-0.105; 0.310]	83.4
PAR	VEN XR	0.030 [-0.192; 0.252]	60.6	0.084 [-0.177; 0.344]	73.6	0.033 [-0.234; 0.300]	59.4
PAR CR	SER	0.072 [-0.179; 0.324]	71.3	-0.064 [-0.314; 0.186]	30.7	0.020 [-0.237; 0.279]	56.3
VEN	PAR CR	0.071 [-0.172; 0.315]	71.7	0.150 [-0.072; 0.373]	90.7	0.124 [-0.108; 0.359]	85.2
VEN XR	PAR CR	0.072 [-0.198; 0.343]	70.0	0.153 [-0.117; 0.421]	86.8	0.195 [-0.093; 0.484]	90.8
VEN	SER	0.142 [-0.062; 0.349]	91.4	0.087 [-0.149; 0.324]	76.3	0.145 [-0.087; 0.376]	89.1
VEN XR	SER	0.144 [-0.093; 0.384]	88.2	0.089 [-0.193; 0.373]	73.2	0.216 [-0.067; 0.500]	93.2
VEN XR	VEN	0.002 [-0.230; 0.232]	50.4	0.003 [-0.255; 0.259]	50.8	0.071 [-0.195; 0.334]	70.1

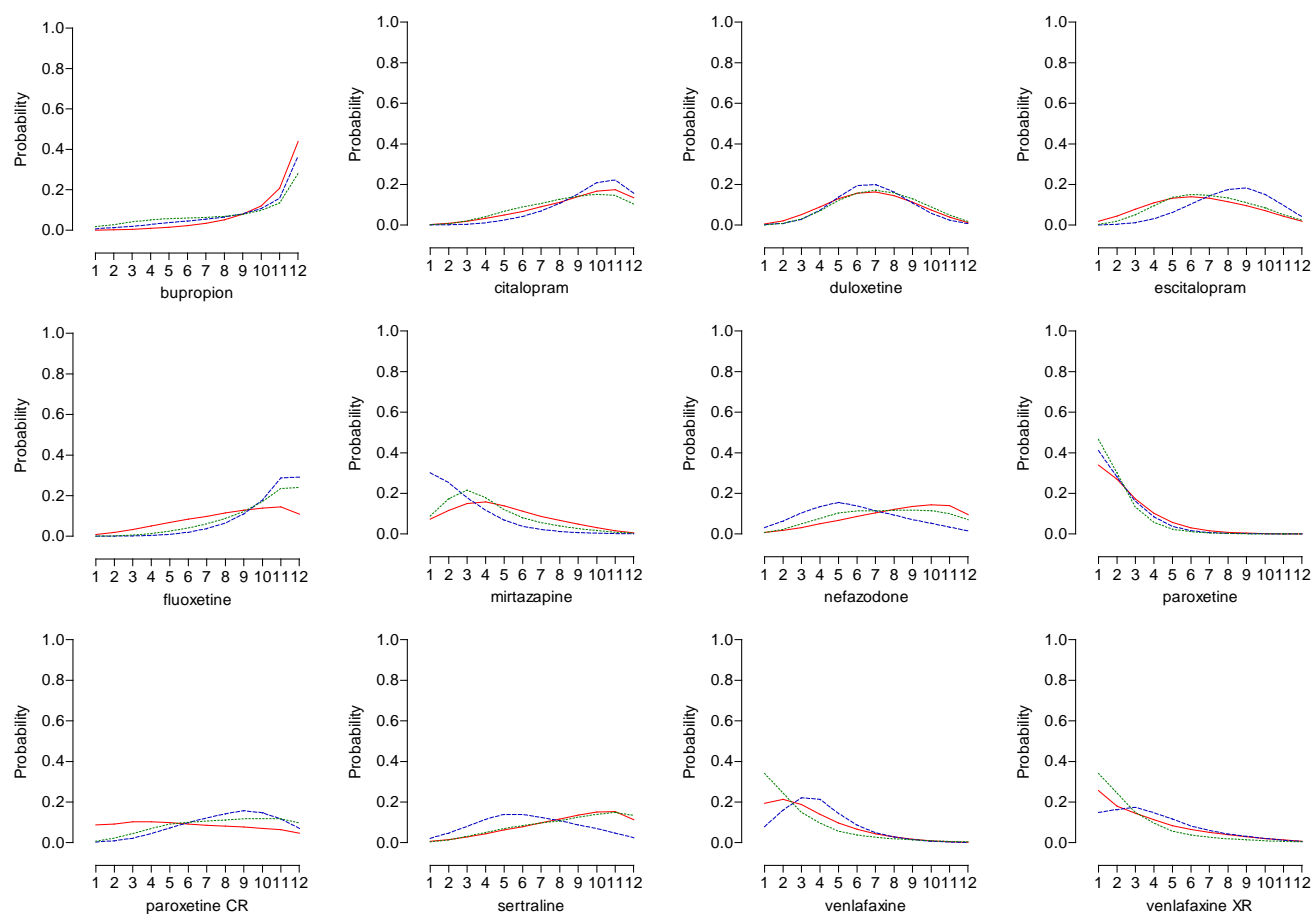
BUP: bupropion; CIT: citalopram; DUL: duloxetine; ESC: escitalopram; FLU: fluoxetine; MIR: mirtazapine; NEF: nefazodone; PAR: paroxetine; PAR CR: paroxetine CR; SER: sertraline; VEN: venlafaxine; VEN XR: venlafaxine XR.

Annexe 1, Figure 1: Graphiques en entonnoir sur données publiées pour les 12 comparaisons entre antidépresseurs et placebo



Sur chaque graphique, le trait vertical représente l'effet combiné et les lignes de niveaux (en pointillés) représentent les seuils de significativité statistique ($p = 0.1$; $p = 0.05$; $p = 0.01$)

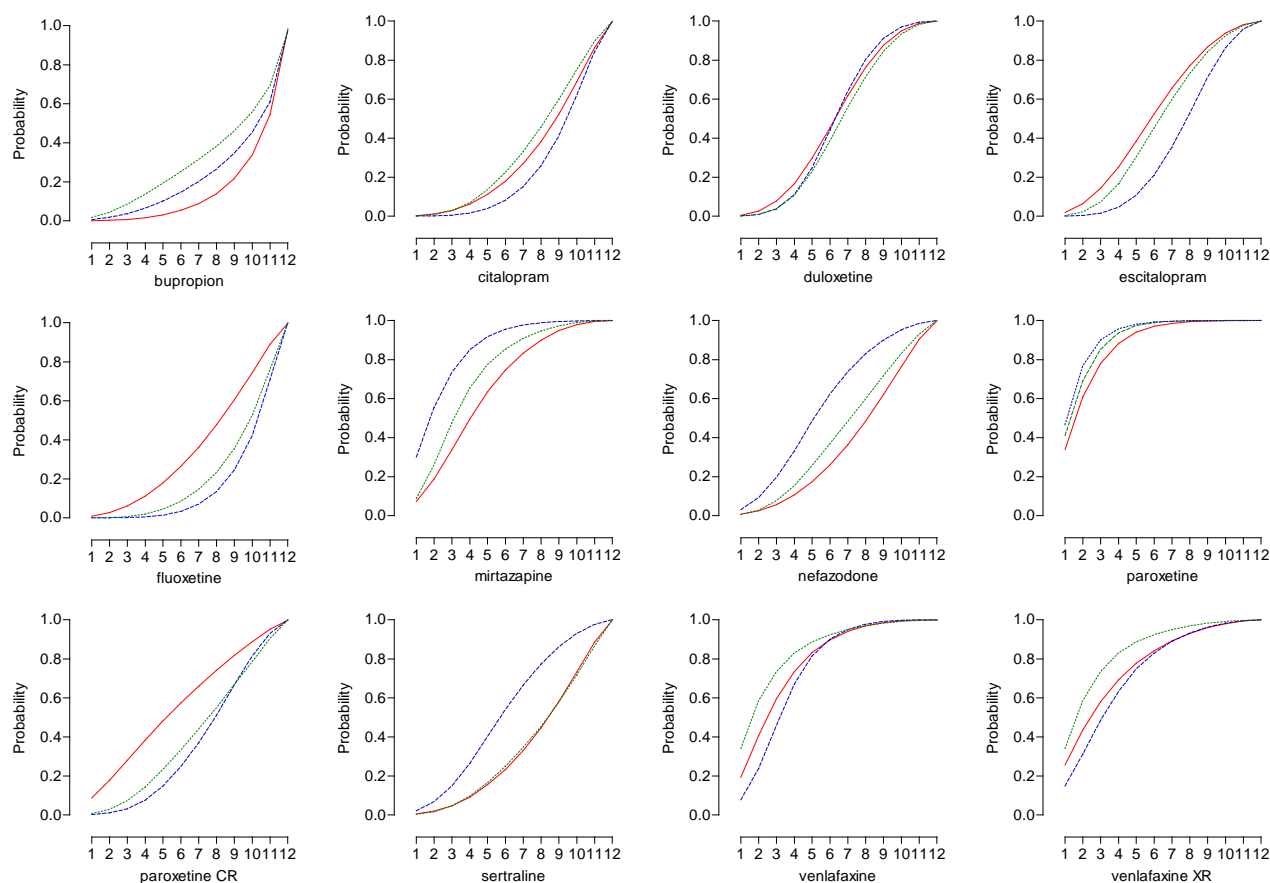
Annexe 1, Figure 2: Courbes des probabilités de rang pour les 12 antidépresseurs



Pour chaque agent, nous avons tracé la probabilité d'être à chaque rang (axe des y) pour chaque rang possible de 1 à 12 (axe des x)

74 essais avec quantités d'effet issues des rapports d'évaluation par la FDA: *trait rouge*; 51 essais avec quantités d'effet issues des articles publiés: *trait bleu*; 51 essais avec quantités d'effet issues des rapports d'évaluation par la FDA: *trait vert*

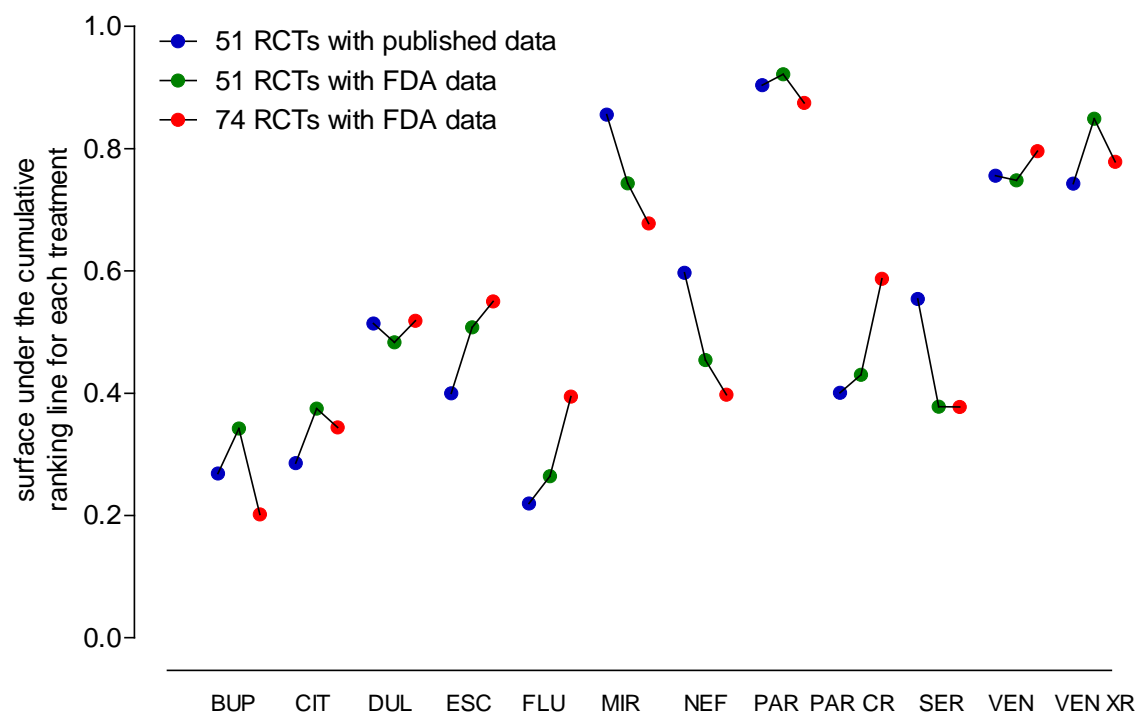
Annexe 1, Figure 3: Courbes des probabilités de rang cumulées pour les 12 antidépresseurs



Pour chaque agent, nous avons tracé la probabilité cumulée d'être le meilleur agent, parmi les 2 meilleurs, parmi les 3 meilleurs, etc. (axe des y) pour chaque rang possible de 1 à 12 (axe des x)

74 essais avec quantités d'effet issues des rapports d'évaluation par la FDA: *trait rouge*; 51 essais avec quantités d'effet issues des articles publiés: *trait bleu*; 51 essais avec quantités d'effet issues des rapports d'évaluation par la FDA: *trait vert*

Annexe 1, Figure 4: Aires sous la courbe des probabilités de rang cumulées pour les 12 antidépresseurs

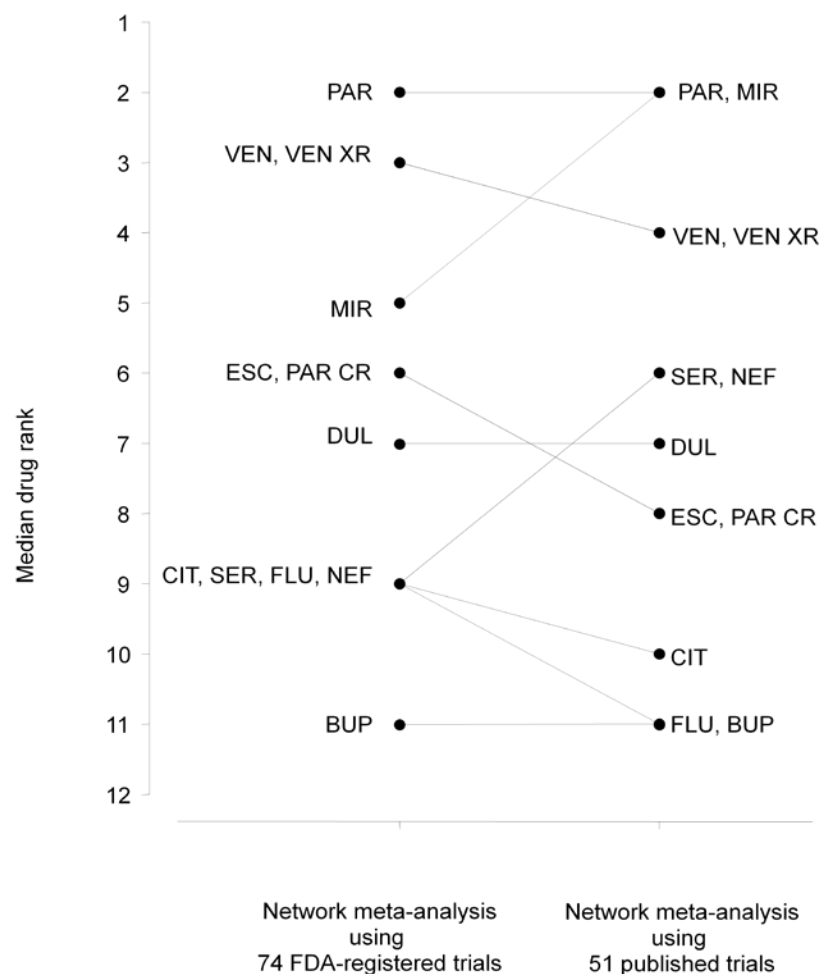


Nous avons représenté l'aire sous la courbe des probabilités de rang cumulées (axe des y) pour chacun des 12 agents (axe des x). Cette aire vaut 1 lorsqu'il est certain que l'intervention considérée est la meilleure et 0 lorsqu'il est certain qu'elle est la pire.

74 essais avec quantités d'effet issues des rapports d'évaluation par la FDA: **point rouge**; 51 essais avec quantités d'effet issues des articles publiés: **point bleu**; 51 essais avec quantités d'effet issues des rapports d'évaluation par la FDA: **point vert**

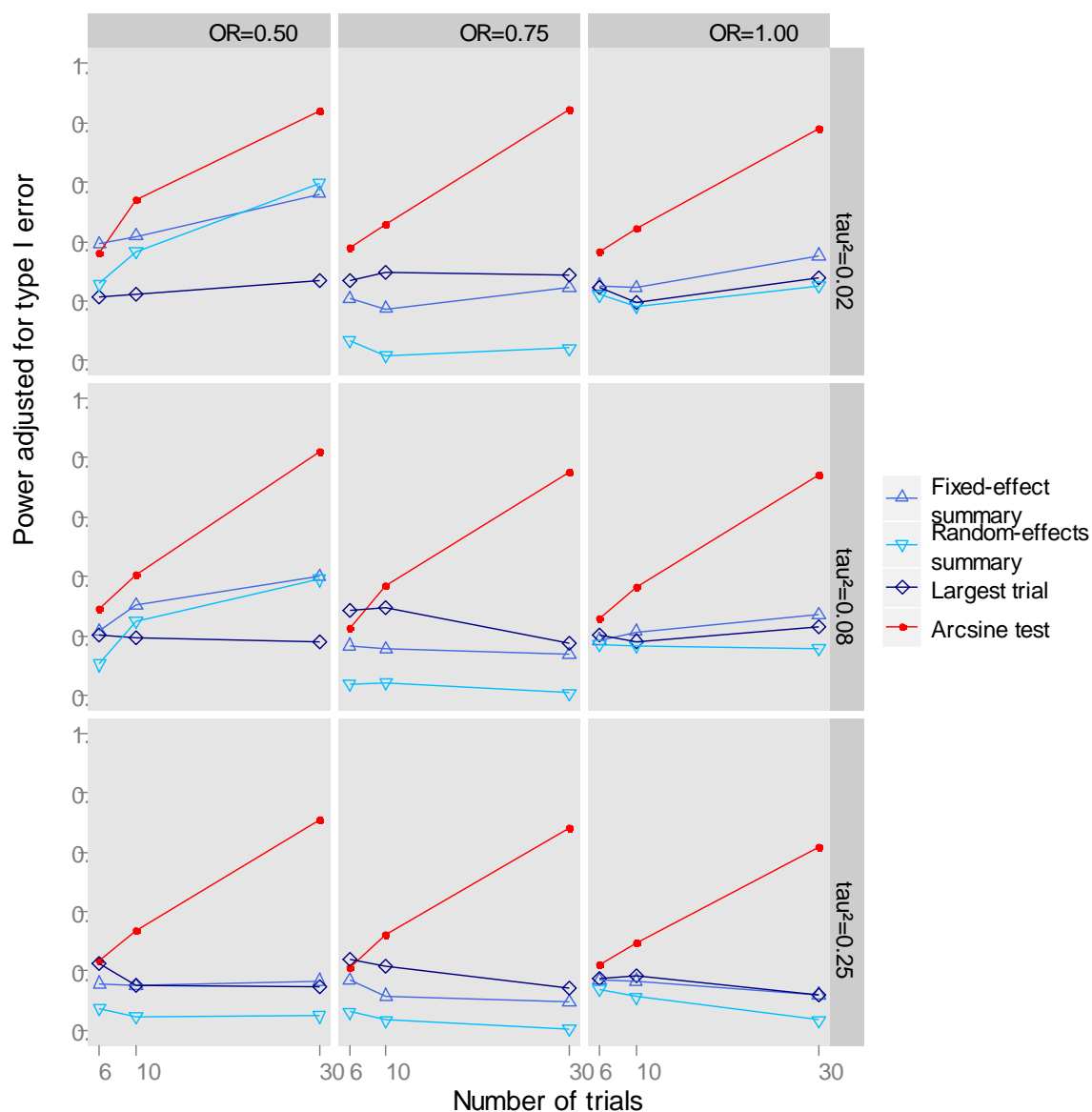
BUP: bupropion; CIT: citalopram; DUL: duloxetine; ESC: escitalopram; FLU: fluoxetine; MIR: mirtazapine; NEF: nefazodone; PAR: paroxetine; PAR CR: paroxetine CR; SER: sertraline; VEN: venlafaxine; VEN XR: venlafaxine XR

Annexe 1, Figure 5: Rangs des 12 antidépresseurs

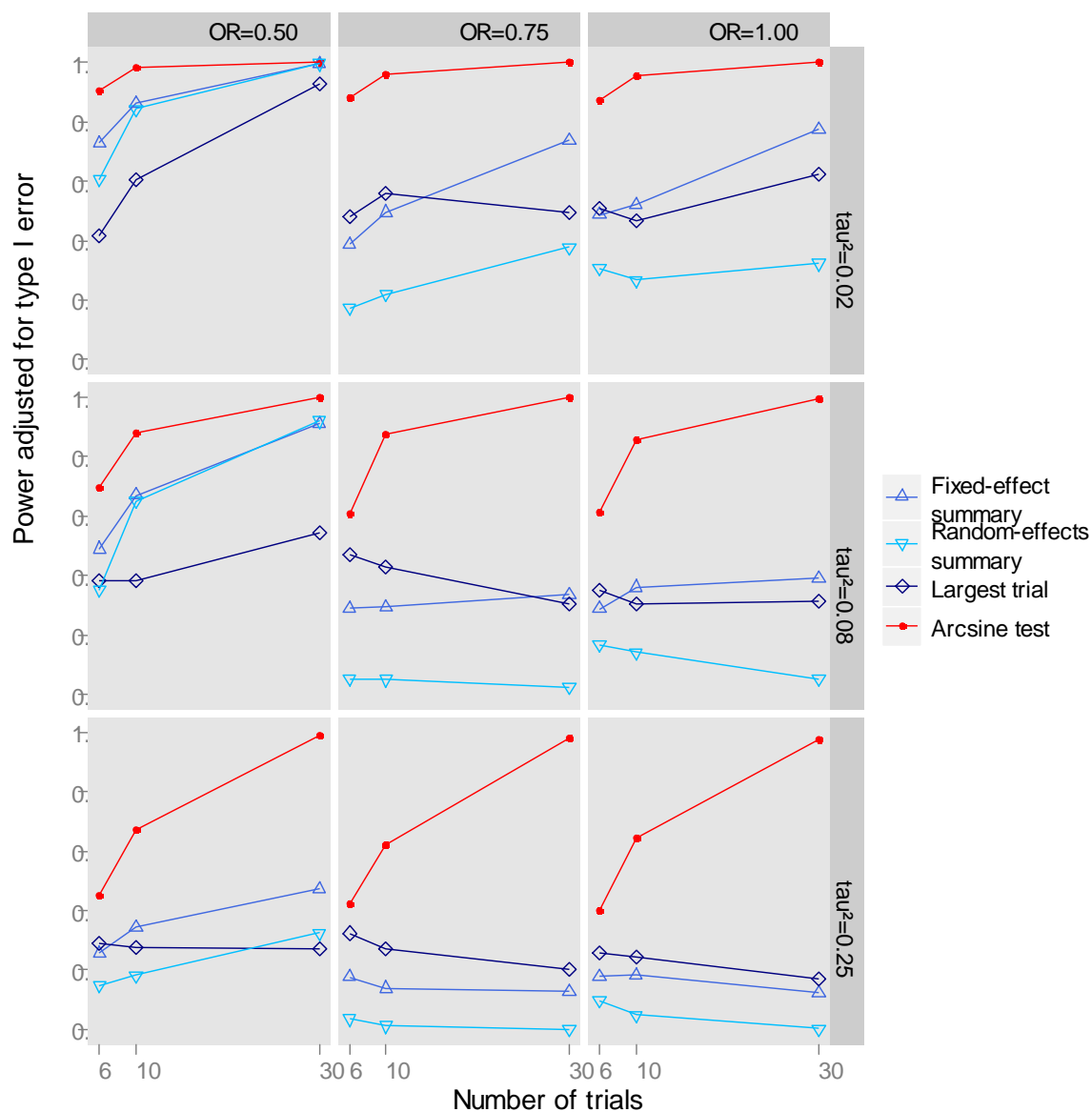


Le classement des 12 agents est basé sur la médiane de la distribution a posteriori des rangs de chaque agent. Via la méta-analyse en réseau des 74 essais enregistrés à la FDA, la paroxétine se classait en première place, la venlafaxine et la venlafaxine XR étaient en deuxième place; via la méta-analyse en réseau des 51 essais avec résultats publiés, la paroxétine et la mirtazapine étaient premiers ex aequo et la venlafaxine et la venlafaxine XR étaient troisièmes ex aequo.

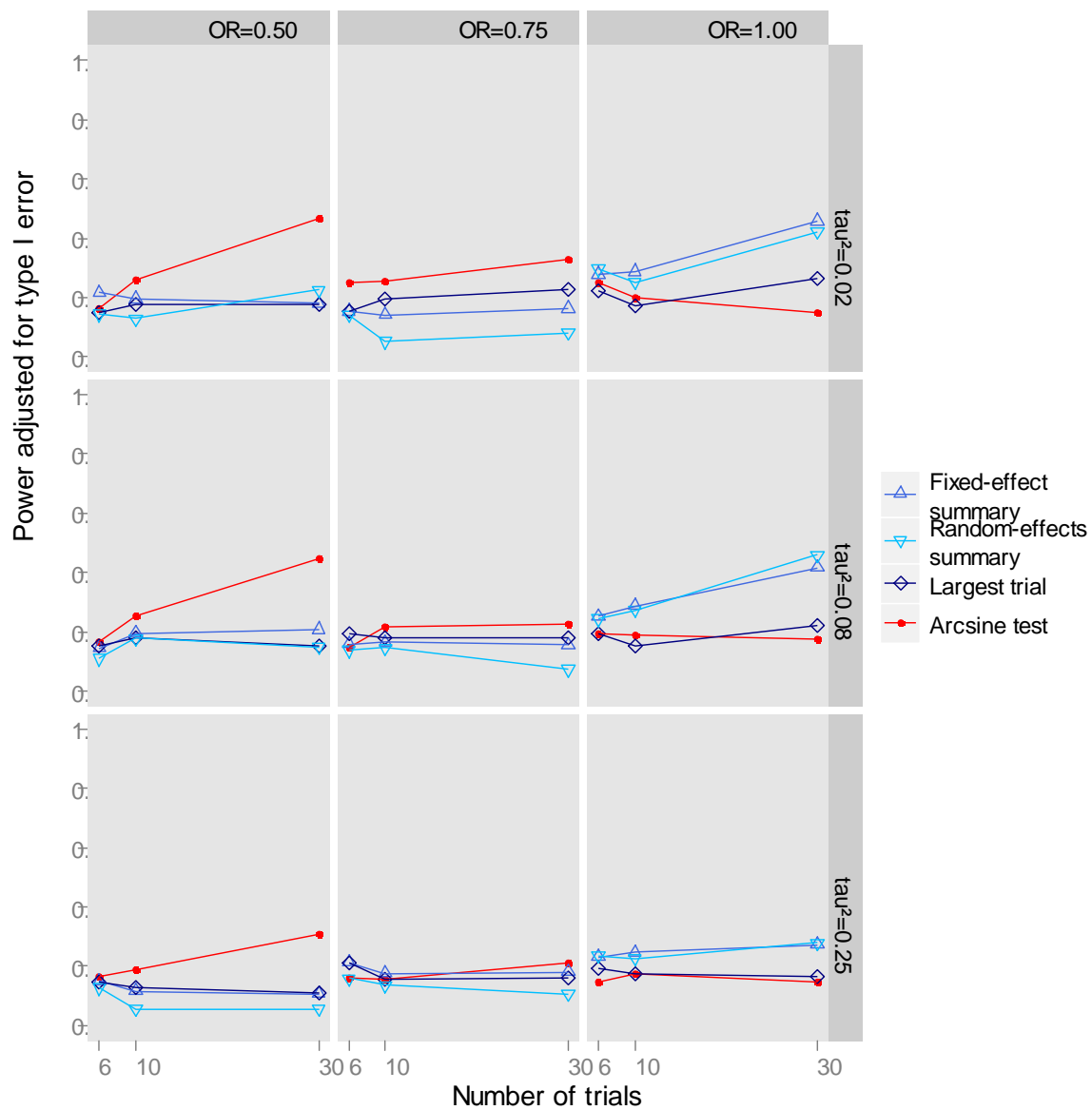
Annexe 2, Figure 1: Puissance des tests pour la méta-analyse conventionnelle (sélection d'essais modélisée en fonction de la taille de l'essai et de l'intensité de l'effet du traitement avec $\rho=0.6$)



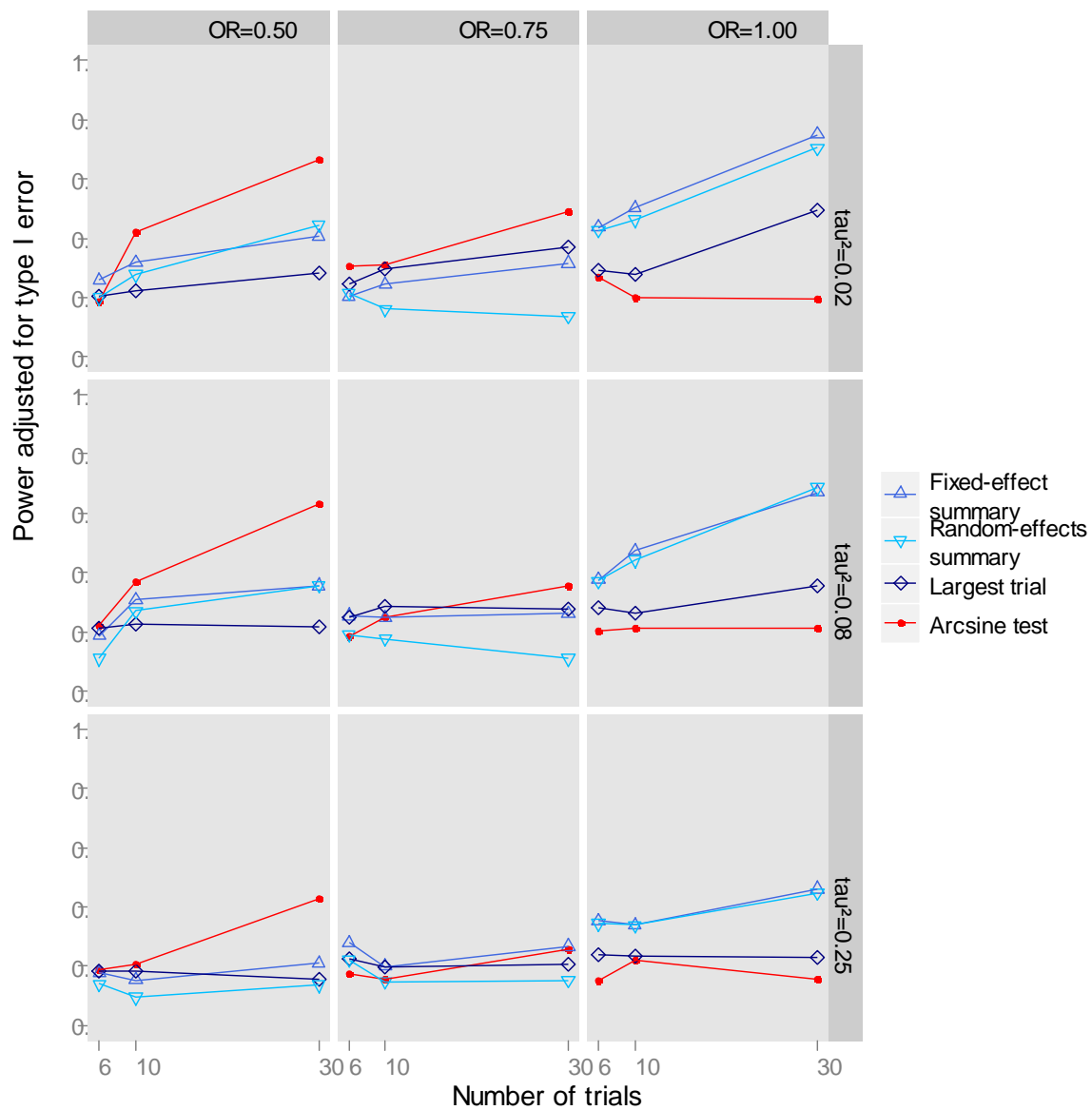
Annexe 2, Figure 2: Puissance des tests pour la méta-analyse conventionnelle (sélection d'essais modélisée en fonction de la taille de l'essai et de l'intensité de l'effet du traitement avec $\rho=1.0$)



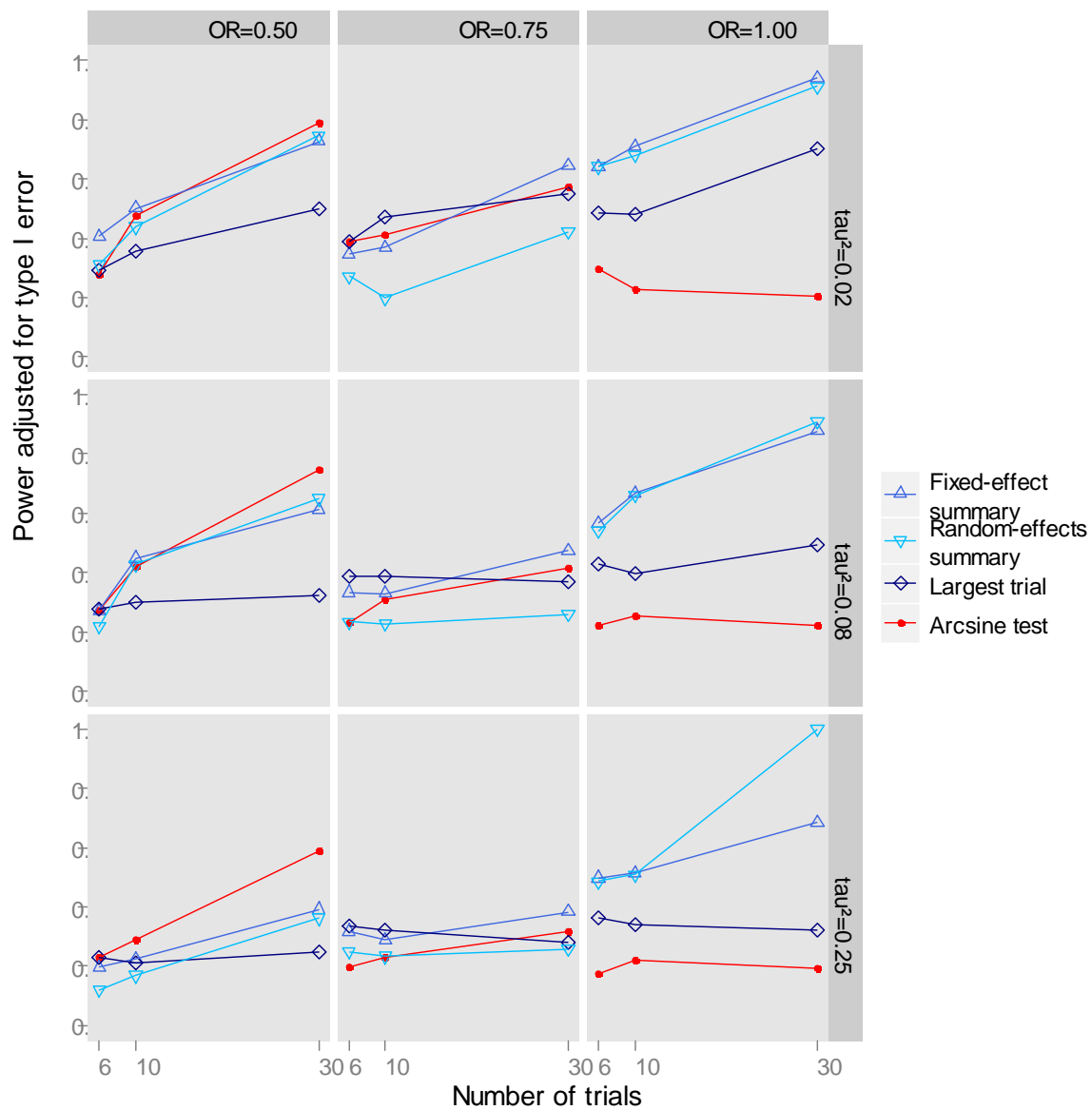
Annexe 2, Figure 3: Puissance des tests pour la méta-analyse conventionnelle (sélection d'essais modélisée en fonction de la valeur p de l'effet du traitement avec $\gamma_0=4$ et $\gamma_1=3$)



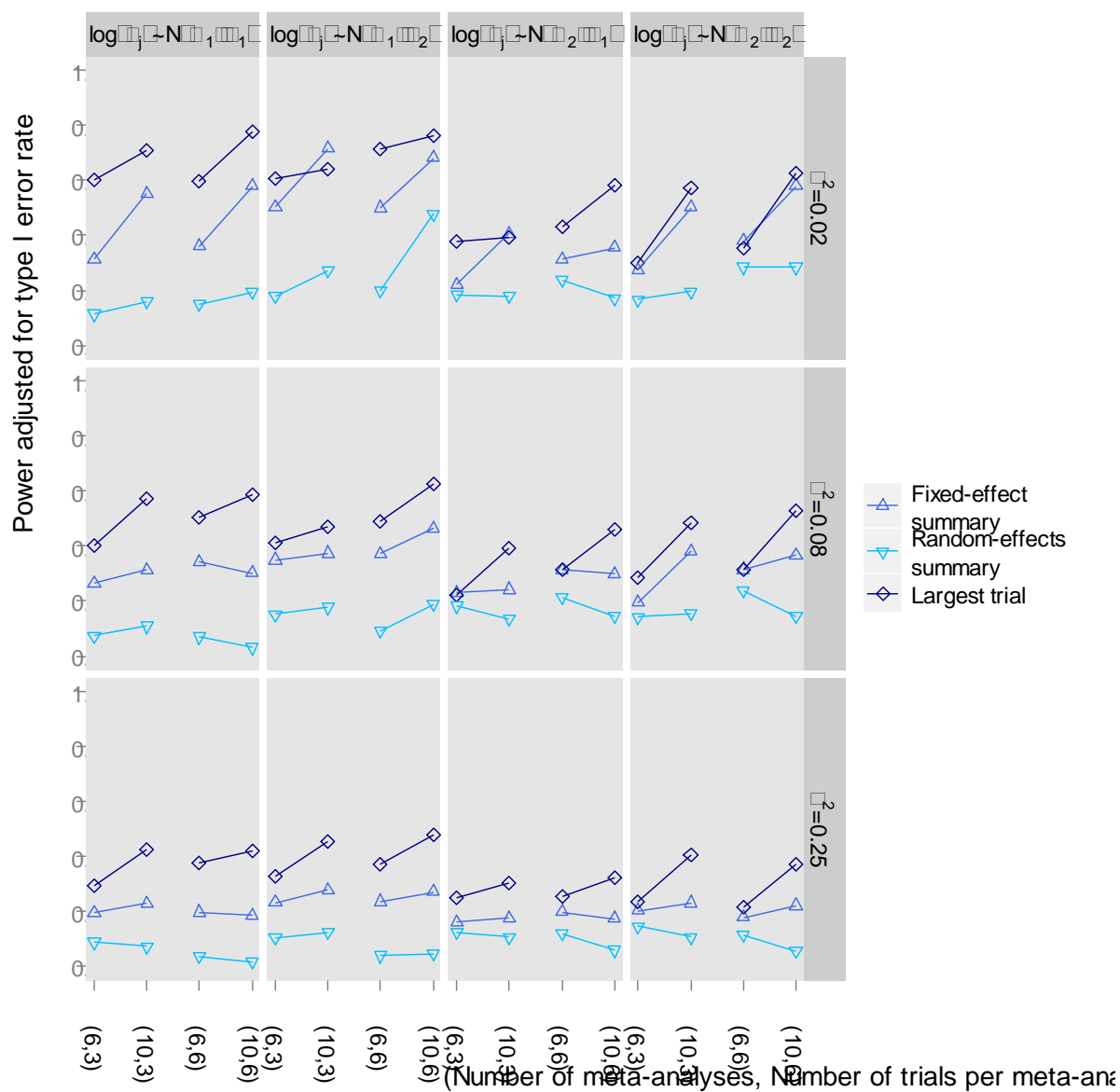
Annexe 2, Figure 4: Puissance des tests pour la méta-analyse conventionnelle (sélection d'essais modélisée en fonction de la valeur p de l'effet du traitement avec $\gamma_0=4$ et $\gamma_1=3/2$)



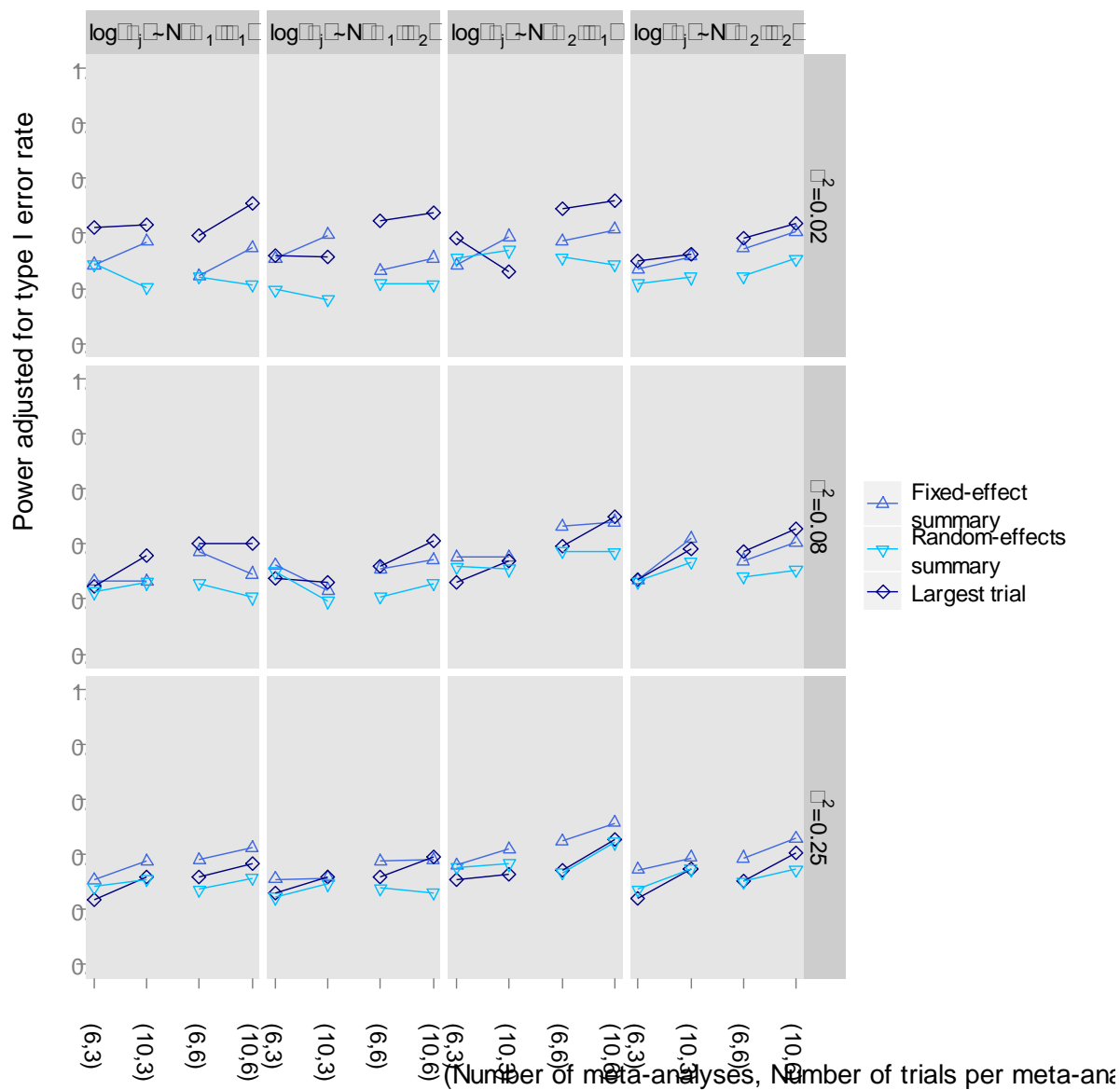
Annexe 2, Figure 5: Puissance des tests pour la méta-analyse conventionnelle (sélection d'essais modélisée en fonction de la valeur p de l'effet du traitement avec $\gamma_0=4$ et $\gamma_1=3/4$)



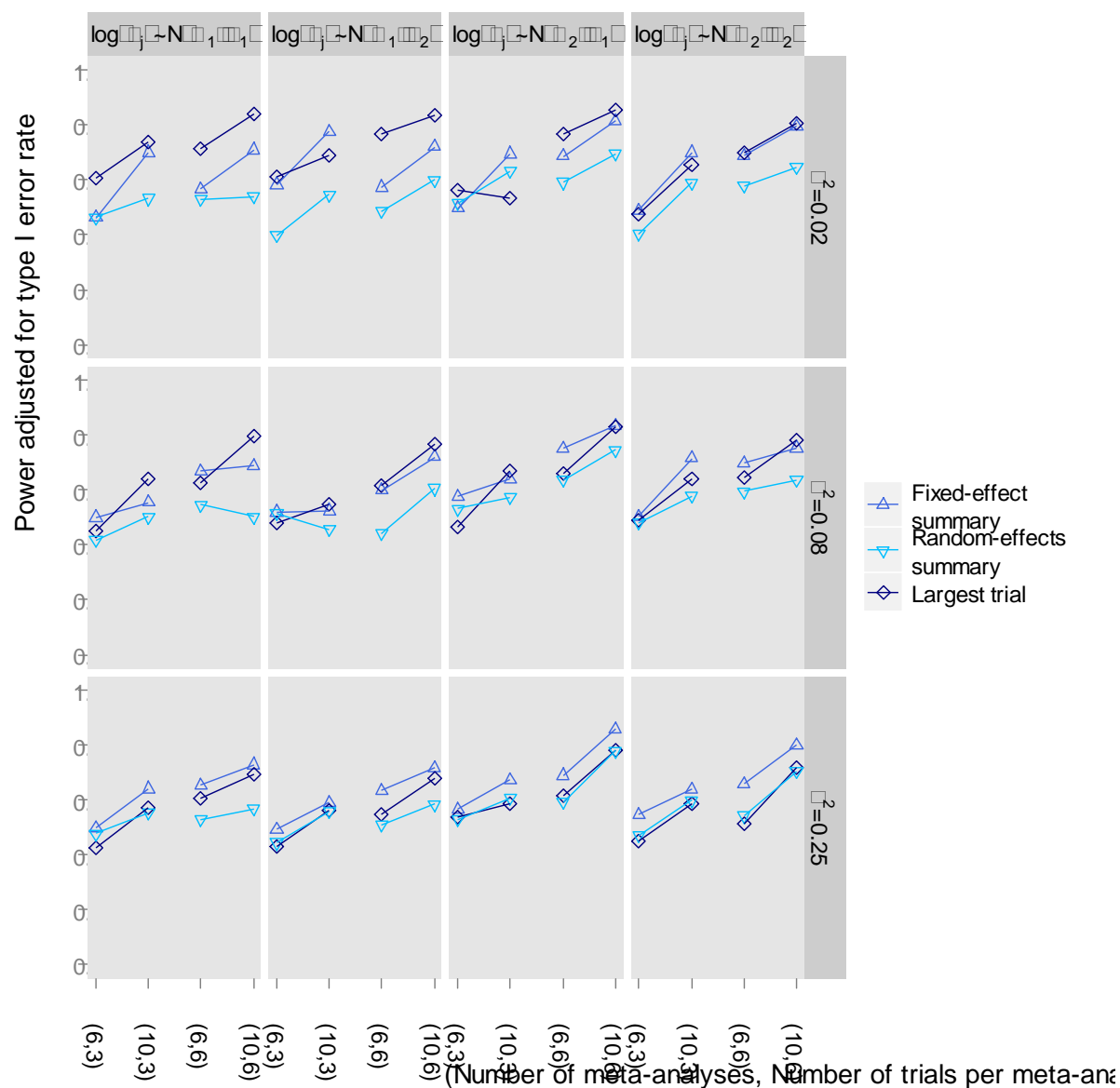
Annexe 2, Figure 6: Puissance des tests pour la méta-analyse en réseau (sélection d'essais modélisée en fonction de la taille de l'essai et de l'intensité de l'effet du traitement avec $\rho_j \sim \text{Uniforme}[-0.8 ; -0.6]$)



Annexe 2, Figure 7: Puissance des tests pour la méta-analyse en réseau (sélection d'essais modélisée en fonction de la valeur p de l'effet du traitement avec $\gamma_{0j}=4$ et $\gamma_{1j}\sim\text{Uniforme}[3/2 ; 3]$)



Annexe 2, Figure 8: Puissance des tests pour la méta-analyse en réseau (sélection d'essais modélisée en fonction de la valeur p de l'effet du traitement avec $\gamma_{0j}=4$ et $\gamma_{1j}\sim\text{Uniforme}[3/4 ; 3/2]$)



Annexe 3, Tableau 1: effets combinés pour les 12 comparaisons entre antidépresseurs et placebo

Agent	N	Données FDA		N	Données publiées	
		SMD (95%CI)	τ^2		SMD (95%CI)	τ^2
Bupropion	3	0.17 [0.04; 0.31]	0.00	1	0.27 [0.01; 0.53]	NA
Citalopram	5	0.25 [0.10; 0.38]	0.00	4	0.30 [0.16; 0.44]	0.00
Duloxetine	8	0.30 [0.21; 0.40]	0.00	6	0.40 [0.29; 0.51]	0.00
Escitalopram	4	0.31 [0.18; 0.44]	0.00	3	0.36 [0.23; 0.48]	0.00
Fluoxetine	5	0.26 [0.06; 0.45]	0.02	5	0.29 [0.01; 0.49]	0.02
Mirtazapine	10	0.35 [0.17; 0.54]	0.04	6	0.57 [0.39; 0.75]	0.00
Nefazodone	6	0.26 [0.12; 0.40]	0.00	4	0.44 [0.26; 0.61]	0.00
Paroxetine	16	0.42 [0.30; 0.54]	0.00	10	0.59 [0.44; 0.74]	0.00
Paroxetine CR	3	0.32 [0.15; 0.49]	0.00	3	0.36 [0.20; 0.51]	0.00
Sertraline	5	0.26 [0.12; 0.39]	0.00	2	0.42 [0.24; 0.60]	0.00
Venlafaxine	6	0.40 [0.24; 0.55]	0.01	5	0.51 [0.36; 0.65]	0.00
Venlafaxine XR	3	0.40 [0.18; 0.62]	0.02	2	0.51 [0.30; 0.71]	0.00

Les quantités d'effet combinées ont été estimées via des modèles de méta-analyse à effets aléatoires. N: nombre d'essais; SMD (IC95%): différence de moyennes standardisée combinée entre chaque agent antidépresseur et placebo (intervalle de confiance à 95%); τ^2 : variance inter-essais; NA: non estimable.

Annexe 3, Tableau 2: Estimations des coefficients des modèles d'ajustement sur données publiées

Modèle de régression		Modèle de sélection			
β_{BUP}	4.487 (4.265)	γ_{0BUP}	3.360 (1.311)	γ_{1BUP}	-10.002 (4.143)
β_{CIT}	4.559 (3.109)	γ_{0CIT}	3.345 (1.299)	γ_{1CIT}	-10.022 (4.128)
β_{DUL}	4.755 (3.726)	γ_{0DUL}	3.346 (1.300)	γ_{1DUL}	-10.011 (4.140)
β_{ESC}	4.570 (4.283)	γ_{0ESC}	3.361 (1.312)	γ_{1ESC}	-10.013 (4.138)
β_{FLU}	4.628 (2.351)	γ_{0FLU}	3.263 (1.284)	γ_{1FLU}	-10.098 (4.083)
β_{MIR}	4.223 (3.496)	γ_{0MIR}	3.521 (1.353)	γ_{1MIR}	-9.883 (4.093)
β_{NEF}	4.761 (3.628)	γ_{0NEF}	3.389 (1.284)	γ_{1NEF}	-10.003 (4.135)
β_{PAR}	3.591 (3.371)	γ_{0PAR}	3.659 (1.411)	γ_{1PAR}	-9.747 (4.052)
$\beta_{PAR\ CR}$	4.507 (4.110)	$\gamma_{0PAR\ CR}$	3.369 (1.306)	$\gamma_{1PAR\ CR}$	-10.008 (4.144)
β_{SER}	4.417 (4.226)	γ_{0SER}	3.374 (1.308)	γ_{1SER}	-10.001 (4.141)
β_{VEN}	5.311 (3.613)	γ_{0VEN}	3.252 (1.289)	γ_{1VEN}	-10.087 (4.143)
$\beta_{VEN\ XR}$	4.536 (4.328)	$\gamma_{0VEN\ XR}$	3.373 (1.317)	$\gamma_{1VEN\ XR}$	-9.996 (4.137)
β	4.532 (2.542)	γ_0	3.237 (1.043)	γ_1	-10.029 (3.977)
σ	2.718 (2.148)	ζ_0	0.731 (0.525)	ζ_1	0.983 (0.574)

Annexe 3, Texte 1: Codes Winbugs

Modèle de régression

```

for (i in 1:50) {
  w[i]<-1/(se[i]*se[i])
  y[i]~dnorm(delta.crude[i],w[i])
  delta.crude[i]<-delta[i] + beta[t[i]]*se[i]*se[i]
  delta[i]~dnorm(mean[i],precision)
  mean[i] <- d[t[i]] - d[b[i]]
}

d[1]<- 0
for (k in 2:13) { d[k] ~ dnorm(0,0.0001)
                  beta[k]~dnorm(beta.mean,precBeta)
}
beta.mean~dnorm(0,0.0001)

sd~dunif(0,10)
precision <- 1/(sd*sd)

sdBeta~dunif(0,10)
precBeta <- 1/(sdBeta*sdBeta)

```

Modèle de sélection

```

for (i in 1:50) { w[i]<-1/(se[i]*se[i])
                  y[i]~dnorm(delta.crude[i],w[i])
                  delta.crude[i]<-delta[i]/weight[i]
                  logit(weight[i])<-gamma0[t[i]]+gamma1[t[i]]*se[i]
                  delta[i]~dnorm(mean[i],precision)
                  mean[i]<- d[t[i]] - d[b[i]]
}

d[1]<- 0
for (k in 2:13) { d[k]~dnorm(0,0.0001)
                  gamma0[k]~dnorm(gamma0.mean,precGamma0)
                  gamma1[k]~dnorm(gamma1.mean,precGamma1)
}

sd~dunif(0,10)
precision<-1/(sd*sd)

p0~dbeta(3.56,4.84)
p1~dbeta(7.52,2.63)

```

```
gamma0.mean<-(logit(p1)/0.098-logit(p0)/0.3180)/(1/0.098-1/0.3180)
```

```
gamma1.mean<-(logit(p1)-logit(p0))/(0.098-0.3180)
```

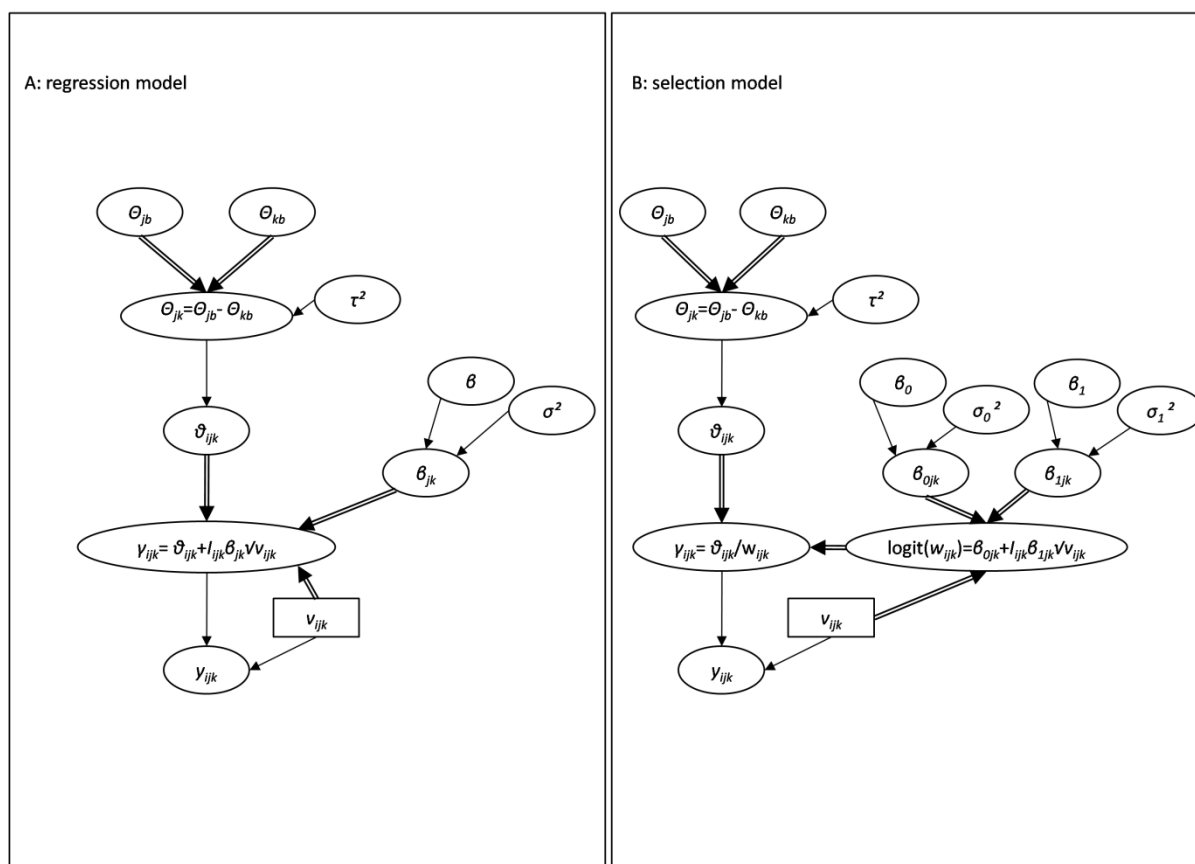
```
sdGamma0~dunif(0,2)
```

```
sdGamma1~dunif(0,2)
```

```
precGamma0<-1/(sdGamma0*sdGamma0)
```

```
precGamma1<-1/(sdGamma1*sdGamma1)
```

Annexe 3, Figure 1: Représentations graphiques des modèles d'ajustement



(A) modèle de régression (B) modèle de sélection